

01-21-00

A

Attorney Docket No.: 5808.200-US

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

FILING UNDER 37 C.F.R. 1.53(b)

Box Patent Application
Assistant Commissioner for Patents
Washington, DC 20231



Express Mail Label No. EL293692871US
Date of Deposit January 20, 2000

Sir:

This is a request for filing a patent application under 37 C.F.R. 1.53(b) of

Applicant(s): Martin Lehmann

Title: Improved Phytases

75 pages of specification 56 sheets of drawings

3 sheets of Declaration and Power of Attorney

[x] The filing fee is calculated as follows:

Basic Fee:	\$ 690.00
Total Claims: $14 - 20 = 0 \times 18 =$	\$0
Independent Claims: $9 - 3 = 6 \times 78 =$	\$ 468.00
Total Fee:	\$1,158.00

Priority of Danish application nos. PA 1999 00092 and PA 1999 01340 filed on January 22, 1999 and September 21, 1999 is claimed under 35 U.S.C. 119. A certified copy is submitted herewith.

Priority of U.S. provisional application nos. 60/117,659 and 60/156,495 filed on January 28, 1999 and September 28, 1999 is claimed under 35 U.S.C. 119.

Please amend the specification as follows: At page 1, after the title, insert
Cross-Reference to Related Applications

This application claims priority of provisional application nos. 60/117,659 and 60/156,495 filed on January 28, 1999 and September 28, 1999 and of Danish application

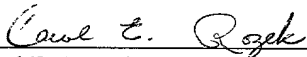
nos. PA 1999 00092 and PA 1999 01340 filed January 22, 1999 and September 21, 1999, the contents of which are fully incorporated herein by reference.

Address all future communications to Steve T. Zelson, Esq., Novo Nordisk of North America, Inc., 405 Lexington Avenue, Suite 6400, New York, NY 10174-6401.

Please charge the required fee, estimated to be \$1,158, to Novo Nordisk of North America, Inc., Deposit Account No. 14-1447. A duplicate of this sheet is enclosed.

Respectfully submitted,

Date: January 20, 2000



Carol E. Rozek, Reg. No. 36,993
Novo Nordisk of North America, Inc.
405 Lexington Avenue, Suite 6400
New York, NY 10174-6401
(212) 867-0123

01-21-00

A

Attorney Docket No.: 5808.200-US

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

EXPRESS MAIL CERTIFICATE

Box Patent Application
Assistant Commissioner for Patents
Washington, DC 20231

Re: U.S. Patent Application for
Title: Improved Phytases
Applicants: Martin Lehmann

Sir:

Express Mail Label No. EL293692871US

Date of Deposit : January 20, 2000

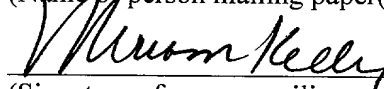
I hereby certify that the following attached paper(s) or fee

1. Filing Under 37 C.F.R. 1.53(b) (in duplicate)
2. Patent Application
3. Unexecuted Combined Declaration and Power of Attorney
4. Certified Copy of Priority Applications

are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, DC 20231.

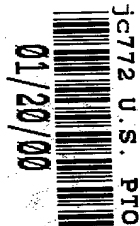
Miriam Kelly

(Name of person mailing paper(s) or fee)



(Signature of person mailing paper(s) or fee)

Mailing Address:
Novo Nordisk of North America, Inc.
405 Lexington Avenue, Suite 6400
New York, NY 10017
(212) 867-0123



09488265 012000

Improved phytases

Phytases are enzymes that hydrolyze phytate (myo-inositol hexakisphosphate) to myo-inositol and inorganic phosphate. They
5 are known to be valuable feed additives.

The present invention relates to improved phytases, viz. phytases of amended characteristics, e.g. amended activity characteristics, reference being made to e.g. the phytase(s) it has been derived from, or to known phytases. Amended activity
10 characteristics means amended in at least one phytase activity related respect, such as (non-exclusive list): pH stability, temperature stability, pH profile, temperature profile, specific activity (in particular in relation to pH and temperature), substrate specificity, substrate cleavage pattern, substrate
15 binding, position specificity, the velocity and level of release of phosphate from corn, reaction rate, phytate degradation rate), end level of released phosphate reached.

Examples of amended activity characteristics are amended specific activity (e.g. increased, e.g. increased at a pH of 3,
20 4, 5, or 6); amended pH or temperature profile; and/or amended (e.g. increased) thermostability, e.g. of an increased melting temperature as measured using Differential Scanning Calorimetry (DSC).

The present invention also relates to a process for the
25 preparation of a modified protein, wherein in a first step a consensus sequence is determined from a number of highly homologous sequences according to steps a), b) and c) below:

a) at least three, preferably at least four amino acid sequences are aligned by any standard alignment program known in
30 the art;

b) at every position of the amino acid sequence alignment, the amino acids are evaluated for their evolutionary similarity and a consensus residue is chosen by any standard program known in the art, whereby the minimal requirements for calculation of a consensus residue are set in such a way that the program is already able to determine a consensus residue if a given residue occurs in only two of the aligned sequences. However, if there is a subgroup of sequences among the compared amino acid sequences that shows a much higher degree of similarity with each other than with the remaining sequences of the alignment, the subgroup may be represented in the calculation only with its consensus sequence determined in the same way as outlined in EP 897985, or alternatively, to each sequence of the subgroup, a vote weight of 1 divided by the number of sequences in the subgroup will be assigned;

c) in case no consensus amino acid at a defined position is identified by the program, any of the amino acids, preferably the most frequently occurring amino acid at this position is selected.

20 In a second aspect of the invention, a homologous sequence is compared with the consensus sequence, and one or more non-consensus residues in this homologous sequence are replaced by the corresponding consensus residues.

Preferably, only such amino acid residues are replaced in
25 the homologous amino acid sequence where a consensus residue can
clearly be defined by the program under moderately stringent
conditions whereas at all positions of the alignment where no
preferred consensus amino acid can be determined under
moderately stringent conditions, the amino acids of the
30 homologous protein remain unchanged.

In a third aspect of the invention, the active center of the protein of interest is determined, comprising all amino acid residues that are involved in forming the active center, both in the consensus sequence, and in the sequence of a homologous protein; subsequently, some or all of the divergent amino acid residues of the homologous protein are inserted in the backbone of the consensus sequence.

In one embodiment of this process, the program used for the comparison of amino acids at a defined position regarding their evolutionary similarity is the program "PRETTY".

The active center of the protein can be determined by using an analysis of the three-dimensional structure of the protein.

An example of a homologous protein is an enzyme family, an example of a defined protein family is the family of phytases, e.g. of fungal origin.

For example, the amino acid sequence of the phytase can be changed by the introduction of at least one mutation or substitution chosen from

20	E58A	F54Y
	D69K	I73V
	D197N	K94A
	T214L	R101A
	E222T	N153K
25	E267D	V158I
	R291I	A203G
	R329H	S205G
	S364T	V217A
	A379K	A227V
30	G404A	V234L
		P238A

000210"5223160

Q277E
 A287H
 A292Q
 V366I
 5 A396S
 E415Q
 G437A
 R451E

For interpreting these abbreviations, as an example, the
 10 mutation E58A is to be interpreted as follows: When subtracting
 26 from the number, you get the position or residue number in
 the consensus phytase sequence or another phytase sequence
 aligned as shown in Fig. 1 (corresponding to the addition of a
 26 amino acid signal sequence to the sequences shown in Fig. 1).
 15 For example, in E58A, number 58 means position number 32 (58-
 26=32). And the letter before the number, i.e. E, represents the
 amino acid in the phytase to be modified which is replaced by
 the amino acid behind the number, i.e. A.

The above-mentioned amino acid replacements, alone and/or
 20 in combination, have a positive effect on the protein stability.

The following sub-groups of mutations are also interesting
 (i.e. phytases comprising at least one mutation selected from
 either one of the groups of):

E58A, D69K, D197N, T214L, E222T, E267D, R291I, R329H,
 25 S364T, A379K, G404A;

F54Y, I73V, K94A, R101A, N153K, V158I, A203G, S205G,
 V217A, A227V, V234L, P238A, Q277E, A287H, A292Q, V366I, A396S,
 E415Q, G437A, R451E;

E58A, D69K, D197N, F54Y, I73V, K94A;
 30 T214L, E222T, E267DR101A, N153K, V158I;
 R291I, R329H, S364TA203G, S205G, V217A;

000210"5222460

A379K, G404AA227V, V234L, P238A, Q277E;
A287H, A292Q, V366I, A396S, E415Q, G437A, R451E;
T214L, E222T, S364T, V158I, A203G, G404A, A227V, P238A,
A396S, G437A, R451E.

5 Examples of host cells are plant cells, animal cells, and
microbial cells, e.g. prokaryotic or eukaryotic cells, such as
bacterial, fungal or yeast cells. An example of a fungal host is
a strain of the genus *Aspergillus*, and examples of yeast hosts
are strains of *Saccharomyces*, and strains of *Hansenula*.

10 The invention also relates to a modified protein
obtainable or obtained by any of the processes described above.

 The invention also relates to a variant or mutein of a
phytase such as (but not limited to) the consensus phytase-1,
wherein, in the amino acid sequence in Figure 2, at least one
15 of the following replacements have been effected: Q50L, Q50T,
Q50G, Q50T-Y51N, Q50L-Y51N or Q50T-K91A.

 In the third aspect mentioned above, a consensus sequence
is determined from homologous sequences as described above; in a
second step the active center of the protein comprising all
20 amino acid residues that are involved in forming the active
center is determined in the consensus sequence and in the
sequence of a single homologous protein as well. The single
homologous protein may have preferred properties like high
specific activity or different pH dependency of enzymatic
25 activity. In a third step some or all amino acid residues that
are involved in forming the active center of the homologous
protein are inserted into the backbone of the consensus
sequence. The result thereof is a chimeric protein having the
active center derived from a single protein and the backbone of
30 the consensus sequence.

The active center of the protein can be determined e.g. by using any analysis of the three-dimensional structure of the protein, e.g. by homology modelling on the basis of a known 3D-structure of a known protein.

5 The present invention also provides consensus proteins obtainable or obtained by such processes, in particular proteins comprising at least one of the amino acid sequences shown in Figures 2-6, 10 or 21, or variants or muteins thereof. Examples of such variants are shown in Figs. 7-9.

10 Such variants or muteins can be defined and prepared on the basis of the teachings given in European Patent Application number 0897010, e.g. Q50L, Q50T, Q50G, Q50L-Y51N, or Q50T-Y51N. These mutations are defined as above, or, alternatively, by reference to Fig. 2. When referring to Fig. 2, no subtraction of
15 the 26 amino acid signal peptide is required (e.g. in "Q50L," at position 50 of the amino acid sequence of Fig. 2, the amino acid Q has been replaced by amino acid L).

A food, feed, or pharmaceutical composition comprising the phytases of the invention is another aspect of the invention.

20 In this context, and relating to the process of the invention, "at least three, preferably at least four amino acid sequences of such defined protein family" means that three, four, five, six to twelve, twenty, fifty, or even more sequences can be used for the alignment and the comparison to create the
25 amino acid sequence of the consensus protein. "Sequences of a defined protein family" means that such sequences fold into a three-dimensional structure, wherein the alpha-helices, the beta-sheets and beta-turns are at the same position so that such structures are, as called by the man skilled in the art, largely
30 superimposable. Furthermore these sequences characterize proteins that show the same type of biological activity, e.g. a

defined enzyme class, e.g. the phytases. The three-dimensional structure of one such protein is sufficient to allow the modelling of the structure of the other homologous proteins of such a family. An example, how this can be done, is given in

5 Example 1. "Evolutionary similarity" in the context of the present invention refers to a scheme which classifies amino acids regarding their structural similarity which allows that one amino acid can be replaced by another amino acid with a minimal influence on the overall structure, as this is done e.g.

10 by programs, like "PRETTY", known in the art. The phrase "the degree of similarity provided by such a program...is set to less stringent number" means in the context of the present invention that values for the parameters which determine the degree of similarity in the program used in the practice of the present

15 invention are chosen in a way to allow the program to define a consensus amino acid for a maximum of positions of the whole amino acid sequence, e. g. in case of the program PRETTY a value of 2 or 3 for the THRESHOLD and a value of 2 for the PLURALITY can be chosen. Furthermore, "a vote weight of one divided by the

20 number of such sequences" means in the context of the present invention that the sequences which define a group of sequences with a higher degree of similarity as the other sequences used for the determination of the consensus sequence only contribute to such determination with a factor which is equal to one

25 divided by the number of all sequences of this group.

As mentioned before, should the program not allow to select the consensus amino acid, the most frequent amino acid is selected; should the latter be impossible the man skilled in the art will select an amino acid from all the sequences used for

30 the comparison which is known in the art for its property to improve the thermostability in proteins as discussed e.g. by

Janecek, S. (1993), *Process Biochem.* 28, 435-445; Fersht, A. R. & Serrano, L. (1993), *Curr. Opin. Struct. Biol.* 3, 75-83; Alber, T. (1989), *Annu. Rev. Biochem.* 58, 765-798; Matthews, B. W. (1987), *Biochemistry* 26, 6885-6888; or Matthews, B. W. (1991),
5 *Curr. Opin. Struct. Biol.* 1, 17-21.

The stability of an enzyme is relevant for many industrial applications. Therefore, a lot of attempts, more or less successful, have been made to improve the stability, preferably the thermostability of enzymes by rational or random approaches.

10 Here we present an alternative way to improve the thermostability of a protein.

The invention provides a process for the preparation of a consensus protein comprising a process to calculate an amino acid residue for nearly all positions of a so-called consensus
15 protein and to synthesize a complete gene from this sequence that can be expressed in a pro- or eukaryotic expression system.

DNA sequences of the present invention can be constructed starting from genomic or cDNA sequences encoding the proteins, e.g. phytases, of interest. For example, they can be constructed
20 by methods of in vitro mutagenesis [see e.g. Sambrook et al., *Molecular Cloning*, Cold Spring Harbor Laboratory Press, New York]. A widely used strategy for "site-directed mutagenesis", as originally outlined by Hurchinson and Edgell [*J. Virol.* 8, 181 (1971)], involves the annealing of a synthetic
25 oligonucleotide carrying the desired nucleotide substitution to a target region of a single-stranded DNA sequence wherein the mutation should be introduced [for review see Smith, *Annu. Rev. Genet.* 19, 423 (1985), and for improved methods, see references 2-6 in Stanssen et al., *Nucl. Acids Res.*, 17, 4441-4454 (1989).
30 Another possibility of mutating a given DNA sequence is the mutagenesis by using the polymerase chain reaction (PCR). DNA as

000210"5323450

starting material can be isolated by methods known in the art and described e.g. in Sambrook et al. (Molecular Cloning) from the respective strains.

For strain information, see e.g. EP 684313 or any
5 depository authority indicated below. *Aspergillus niger* [ATCC 9142], *Myceliophthora thermophila* [ATCC 48102], *Talaromyces thermophilus* [ATCC 20186] and *Aspergillus fumigatus* [ATCC 34625] have been redeposited according to the conditions of the Budapest Treaty at the American Type Culture Cell Collection
10 under the following accession numbers: ATCC 74337, ATCC 74340, ATCC 74338 and ATCC 74339, respectively. It is, however, understood that DNA encoding a consensus protein in accordance with the present invention can also be prepared in a synthetic manner as described, e.g. in EP 747483 or EP 897985, or in the
15 examples, by methods known in the art.

For sequence information, see e.g. EP 684313, or sequence data bases, for example like Genbank (Intelligenetics, California, USA), European Bioinformatics Institute (Hinton Hall, Cambridge, GB), NBRF (Georgetown University, Medical
20 Centre, Washington DC, USA) and Vecbase (University of Wisconsin, Biotechnology Centre, Madison, Wisconsin, USA).

The process of the present invention can e.g. be used to improve the thermostability of the enzyme phytase.

Once complete DNA sequences of the present invention have
25 been obtained they can be integrated into vectors by methods known in the art and described e.g. in Sambrook et al. (s.a.) to overexpress the encoded polypeptide in appropriate host systems. However, a man skilled in the art knows that also the DNA sequences themselves can be used to transform the suitable host
30 systems of the invention to get overexpression of the encoded polypeptide. Appropriate host systems are for example fungi,

000210"59232150

like Aspergilli, e.g. *Aspergillus niger* [ATCC 9142] or
Aspergillus ficuum [NRRL 3135] or like *Trichoderma*, e.g.
Trichoderma reesei; or yeasts, like *Saccharomyces*, e.g.
Saccharomyces cerevisiae or *Pichia*, like *Pichia pastoris*, or
5 *Hansenula polymorpha*, e.g. *H. polymorpha* (DSM5215); or plants,
as described, e.g. by Pen et al., *Bio/Technology* 11, 811-814
(1994). A man skilled in the art knows that such microorganisms
are available from depository authorities, e.g. the American
Type Culture Collection (ATCC), the Centraalbureau voor
10 Schimmelcultures (CBS) or the Deutsche Sammlung für
Mikroorganismen und Zellkulturen GmbH (DSM) or any other
depository authority as listed in the Journal "Industrial
Property" [(1991) 1, pages 29-40]. Bacteria which can be used
are e.g. *E. coli*; *Bacilli* as, e.g., *Bacillus subtilis*; or
15 *Streptomyces*, e.g. *Streptomyces lividans* (see e.g. Anné and
Mallaert in *FEMS Microbiol. Lett.* 114, 121 (1993). Preferred *E.*
coli strains, which can be used are *E. coli* K12 strains e.g. M15
[described as DZ 291 by Villarejo et al. in *J. Bacteriol.* 120,
466-474 (1974)], HB 101 [ATCC No. 33694] or *E. coli* SG13009
20 [Gottesman et al., *J. Bacteriol.* 148, 265-273 (1981)].

Vectors which can be used for expression in fungi are
known in the art and described e.g. in EP 420358, or by Cullen
et al. [*Bio/Technology* 5, 369-376 (1987)], Ward [*Molecular*
Industrial Mycology, Systems and Applications for Filamentous
25 *Fungi*, Marcel Dekker, New York (1991)], Upshall et al.
[*Bio/Technology* 5, 1301-1304 (1987)], Gwynne et al.
[*Bio/Technology* 5, 71-79 (1987)], or Punt et al. [*J. Biotechnol.*
17, 19-34 (1991)]; and for yeasts by Sreekrishna et al. [*J.*
Basic Microbiol. 28, 265-278 (1988), *Biochemistry* 28, 4117-4125
30 (1989)], Hitzemann et al. [*Nature* 293, 717-722 (1981)] or in EP
183070, EP 183071, EP 248227, or EP 263311. Suitable vectors

which can be used for expression in *E. coli* are mentioned, e.g. by Sambrook et al. [s.a.], Fiers et al. [Procd. 8th Int. Biotechnology Symposium", Soc. Franc. de Microbiol., Paris (Durand et al., eds.), pp. 680-697 (1988)], Bujard et al. [Meth. 5 Enzymol. 155, 416-433 (1987)], or Stüber et al. [Immunological Methods, eds. Lefkovits and Pernis, Academic Press, Inc., Vol. IV, 121-152 (1990)]. Vectors that can be used for expression in *Bacilli* are known in the art and described, e.g. in EP 207459, EP 405370, Proc. Natl. Acad. Sci. USA 81, 439 (1984) or Yansura 10 and Henner, Meth. Enzymol. 185, 199-228 (1990). Vectors which can be used for the expression in *H. Polymorpha* are known in the art and described, e.g. in Gellissen et al., Biotechnology 9, 291-295 (1991).

Either such vectors already carry regulatory elements, 15 e.g. promoters, or the DNA sequences of the present invention can be engineered to contain such elements. Suitable promotor elements which can be used are known in the art and are, e.g. for *Trichoderma reesei* the *cbh1*- [Haarki et al., Biotechnology 7, 596-600 (1989)] or the *pk11*-promotor [Schindler et al., Gene 20 130, 271-275 (1993)]; for *Aspergillus oryzae* the *amy*-promotor [Christensen et al., Abstr. 19th Lunteren Lectures on Molecular Genetics F23 (1987), Christensen et al., Biotechnology 6, 1419-1422 (1988), Tada et al., Mol. Gen. Genet. 229, 301 (1991)]; and for *Aspergillus niger* the *glaA*- [Cullen et al., Bio/Technology 25 5, 369-376 (1987), Gwynne et al., Bio/Technology 5, 713-719 (1987), Ward in Molecular Industrial Mycology, Systems and Applications for Filamentous Fungi, Marcel Dekker, New York, 83-106 (1991)], *alcA*- [Gwynne et al., Bio/Technology 5, 718-719 (1987)], *suc1*- [Boddy et al., Curr. Genet. 24, 60-66 (1993)], 30 *aphA*- [MacRae et al., Gene 71, 339-348 (1988), MacRae et al., Gene 132, 193-198 (1993)], *tpiA*- [McKnight et al., Cell 46, 143-

147 (1986), Upshall et al., Bio/Technology 5, 1301-1304 (1987)],
gpdA- [Punt et al., Gene 69, 49-57 (1988), Punt et al., J.
Biotechnol. 17, 19-37 (1991)] and the pkiA-promotor [de Graaff
et al., Curr. Genet. 22, 21-27 (1992)]. Suitable promotor
5 elements that can be used for expression in yeast are known in
the art and are, e.g. the pho5-promotor [Vogel et al., Mol.
Cell. Biol., 2050-2057 (1989); Rudolf and Hinnen, Proc. Natl.
Acad. Sci. 84, 1340-1344 (1987)] or the gap-promotor for
expression in *Saccharomyces cerevisiae*; the aox1-promotor [Koutz
10 et al., Yeast 5, 167-177 (1989); Sreekrishna et al., J. Basic
Microbiol. 28, 265-278 (1988)] for *Pichia pastoris*; or the FMD
promoter [Hollenberg et al., EPA No. 0299108] or MOX-promotor
[Ledeboer et al., Nucl. Acids Res. 13, 3063-3082 (1985)] for *H.*
polymorpha.

15 Accordingly vectors comprising DNA sequences of the
present invention, preferably for the expression of said DNA
sequences in bacteria or a fungal or a yeast host and such
transformed bacteria or fungal or yeast hosts are also a part of
the invention.

20 The invention also provides a system that allows for high
expression of proteins, in particular of the phytases of the
invention, such as recombinant *Hansenula* strains. To achieve
that, the codons of the DNA sequence of such a protein may be
selected on the basis of a codon frequency table of the organism
25 used for expression, e.g. of yeast as in the present case (see
e.g. in Example 1). Optionally, the codons for the signal
sequence may be selected in a manner as described for the
specific case in Example 1; that means that a codon frequency
table is prepared on the basis of the codons used in the DNA
30 sequences which encode the amino acid sequences of the given
protein family. Then the codons for the design of the DNA

sequence of the signal sequence are selected from a codon frequency table of the host cell used for expression whereby always codons of comparable frequency in both tables are used.

Once such DNA sequences have been expressed in an appropriate host cell in a suitable medium, the encoded protein can be isolated either from the medium in the case the protein is secreted into the medium or from the host organism in case such protein is present intracellularly by methods known in the art of protein purification or described in case of a phytase, e.g. in EP 420358. Accordingly, a process for the preparation of a polypeptide of the present invention wherein transformed bacteria or a host cell as described above are cultured under suitable culture conditions, and the polypeptide is recovered therefrom and a polypeptide when produced by such a process; or a polypeptide encoded by a DNA sequence of the present invention, are also a part of the present invention.

Once obtained, the polypeptides of the present invention can be characterized regarding their properties that make them useful in agriculture by any assay known in the art.

In general, the polypeptides of the present invention can be used without being limited to a specific field of application, e.g. in case of phytases for the conversion of inositol polyphosphates, like phytate, to inositol and inorganic phosphate.

Furthermore, the polypeptides of the present invention can be used in a process for the preparation of a pharmaceutical composition or compound food or feeds wherein the components of such a composition are mixed with at least one polypeptide of the present invention. Accordingly, compound food or feeds or pharmaceutical compositions comprising at least one polypeptide of the present invention are also a part of the present

invention. A man skilled in the art is familiar with their process of preparation. Such pharmaceutical compositions or compound foods or feeds can further comprise additives or components generally used for such purpose and known in the state of the art.

The present invention also provides a process for the reduction of levels of phytate in animal manure wherein an animal is fed such a feed composition in an amount effective in converting phytate contained in the feedstuff to lower inositol phosphates and/or inositol, and inorganic phosphate.

In the present context, a phytase is an enzyme or polypeptide that has phytase activity. A phytase can be e.g. a myo-inositol hexakisphosphate phosphohydrolase, such as (myo-inositol hexakisphosphate 3-phosphohydrolase, EC 3.1.3.8) and (myo-inositol hexakisphosphate 6-phosphohydrolase, EC 3.1.3.26).

In one embodiment, the phytase is purified, viz. at least 85%, preferably at least 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% pure, as evaluated by SDS-PAGE. The phytase may be isolated. Phytase activity can be determined using any phytase assay known in the art, e.g. the assay described herein (see Example 9). The assay temperature may be the optimum temperature of the actual phytase, and the assay pH may be the optimum pH of the actual phytase.

The assay temperature may e.g. be selected within the range of 20-90°C, or 30-80°C, or 35-75°C, for instance temperatures of 37°C, 50°C, 60°C, or 70°C.

The assay pH may e.g. be selected within the range of pH 2-9, or 3-8, or 3-6, for instance assay pH values of 3, 4, 5, 6, or 7 may be chosen.

Amino acid sequence homology (or polypeptide or amino acid homology) is determined as the degree of identity between two

sequences. This may suitably be determined by means of computer programs known in the art such as GAP provided in the GCG program package [Program Manual for the Wisconsin Package, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin 53711, USA], see also Needleman, S.B. and Wunsch, C.D., (1970), J. Mol. Biol., 48, 443-453]. In release 9.1, for comparing polypeptide sequences, the Length Weight is set to 0, and the Gap Weight is set to 3.0.

The degree of identity or homology between two DNA (nucleic acid) sequences may be determined by means of computer programs known in the art such as GAP provided in the GCG program package [Program Manual for the Wisconsin Package, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin 53711, USA), see also Needleman, S.B. and Wunsch, C.D., (1970), J. Mol. Biol., 48, 443-453]. In release 9.1, GAP is used with the following settings for DNA sequence comparison: GAP creation penalty of 50 and GAP extension penalty of 3.

Suitable experimental conditions for determining whether a given DNA or RNA sequence hybridizes to a specified nucleotide or oligonucleotide probe involves presoaking of the filter containing the DNA or RNA fragments to examine for hybridization in 5 x SSC (Sodium chloride/Sodium citrate; (J. Sambrook, E.F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2nd edition, Cold Spring Harbor, New York) for 10 min, and prehybridization of the filter in a solution of 5 x SSC, 5 x Denhardt's solution, 0.5 % SDS and 100 µg/ml of denatured sonicated salmon sperm DNA (Sambrook et al. 1989), followed by hybridization in the same solution containing a concentration of 10 ng/ml of a random-primed (Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. 132:6-13), ³²P-dCTP-labeled (specific

activity $> 1 \times 10^9$ cpm/ μ g) probe for 12 hours at approximately 45°C.

The filter is then washed twice for 30 minutes in 2 x SSC, 0.5 % SDS at at least 55°C (low stringency), at at least 60°C (medium stringency), at at least 65°C (medium/high stringency), at at least 70°C (high stringency), or at at least 75°C (very high stringency).

Molecules to which the oligonucleotide probe hybridizes under these conditions can be detected using an x-ray film.

Phytases of amended thermostability, or thermostable phytases, are one aspect of the present invention. A "thermostable" phytase is a phytase that has a T_m (melting temperature) - as measured on purified phytase protein by Differential Scanning Calorimetry (DSC) - of at least 65°C. For the DSC, a constant heating rate may be used, e.g. of 10°C/min. In alternative embodiments, the T_m is at least 66, 67, 68, 69, 70, 71, 72, 73, 74 or 75°C. Or, the T_m is equal to or lower than 150°C, or equal to or lower than 145, 140, 135, 130, 125, 120, 115 or 110°C. Accordingly, examples of intervals of T_m are: 65-150°C, 66-150°C, - (etc.) - 75-150°C; 65-145°C, 66-145°C, - (etc.) - 75-145°C; 65-140°C, - (etc.) - 75-140°C; - (etc.) - 65-110°C, 66-110°C, - (etc.) - 75-110°C.

Particular ranges for T_m are the following: between 65 and 110°C; between 70 and 110°C; between 70 and 100°C; between 75 and 95°C, or between 80 and 90°C.

In Examples 9 and 10 below, the measurement of T_m by DSC is described, and the T_m 's of a number of phytases are shown.

The optimum temperatures are also indicated, since - as an alternative mean - a thermostable phytase can be defined as a phytase having a temperature-optimum of at least 60°C. Preferably, the optimum temperature is determined on the

substrate phytate or phytic acid at pH 5.0 or 5.5. Example 9 describes an example of a phytase assay, including a definition of units.

In alternative embodiments, the optimum temperature is at least 61, 62, 63, 64, 65, 66, 67, 68, 69 or 70°C. In a particular embodiment, the optimum temperature is equal to or lower than 140°C, or equal to or lower than 135, 130, 125, 120, 115, 110, 105 or 100°C. Accordingly, examples of intervals of optimum temperature are: 60-140°C, 61-140°C, - (etc.) - 70-140°C; 60-135°C, 61-135°C, - (etc.) - 70-135°C; 60-130°C, - (etc.) - 70-130°C; - (etc.) - 60-100°C, 61-100°C, - (etc.) - 70-100°C.

Before describing the present invention in more detail a short explanation of the Figures enclosed is given below.

15

Figure 1: Design of the consensus phytase-1 sequence. The letters represent the amino acid residues in the one-letter code. The following sequences were used for the alignment: phyA from *Aspergillus terreus* 9A-1 [Mitchell, D. B., Vogel, K., Weimann, B. J., Pasamontes, L. & van Loon, A. P. G. M. (1997) The phytase subfamily of histidine acid phosphatases: isolation of genes for two novel phytases from the fungi *Aspergillus terreus* and *Myceliophthora thermophila*, *Microbiology* 143, 245-252); from amino acid (aa) 27; SEQ ID NO: 1]; phyA from *A. terreus* cbs116.46 [EP 897985]. A heat resistant phytase of *Aspergillus fumigatus* with superior performance in animal experiments. Phytase optimization and natural variability. In: *The Biochemistry of phytate and phytases* (eds. Rasmussen, S.K; Raboy, V.; Dalbøge, H. and Loewus, F.; Kluwer Academic Publishers); from aa 27; SEQ ID NO: 2; phyA from *Aspergillus niger* var. *awamori* (Piddington et al (1993) *Gene* 133, 55-62;

from aa 27; SEQ ID NO: 3); phyA from *A. niger* T213 (EP 897985);
from aa 27; SEQ ID NO: 4); phyA from *A. niger* strain NRRL3135
[van Hartingsveldt, W., van Zeijl, C. M. F., Hartevelde, G. M.,
Gouka, R. J., Suykerbuyk, M. E. G., Luiten, R. G. M., van
5 Paridon, P. A., Selten, G. C. M., Veenstra, A. E., van Gorcom,
R. F. M., & van den Hondel, C. A. M. J. J. (1993) Cloning,
characterization and overexpression of the phytase-encoding gene
(phyA) of *Aspergillus niger*. *Gene* 127, 87-94; from aa 27; SEQ ID
NO: 5]; phyA from *Aspergillus fumigatus* ATCC 13073 (Pasamontes,
10 L., Haiker, M., Wyss, M., Tessier, M. & van Loon, A. P. G. M.
(1997) Cloning, purification and characterization of a heat
stable phytase from the fungus *Aspergillus fumigatus*, *Appl.*
Environ. Microbiol. 63, 1696-1700; from aa 25; SEQ ID NO: 6];
phyA from *A. fumigatus* ATCC 32722 (EP 897985); from aa 27; SEQ
15 ID NO: 7); phyA from *A. fumigatus* ATCC 58128 (EP 897985); from
aa 27; SEQ ID NO: 8); phyA from *A. fumigatus* ATCC 26906
(EP 897985); from aa 27; SEQ ID NO: 9); phyA from *A. fumigatus*
ATCC 32239 (EP 897985); from aa 30; SEQ ID NO: 10; phyA from
Emericella nidulans [Pasamontes, L., Haiker, M., Henriquez-
20 Huecas, M., Mitchell, D. B. & van Loon, A. P. G. M. (1997a).
Cloning of the phytases from *Emericella nidulans* and the
thermophilic fungus *Talaromyces thermophilus*. *Biochim. Biophys.*
Acta 1353, 217-223; from aa 25; SEQ ID NO: 11]; phyA from
Talaromyces thermophilus (Pasamontes et al., 1997a; from aa 24;
25 SEQ ID NO: 12); and phyA from *Myceliophthora thermophila*
(Mitchell et al., 1997; from aa 19; SEQ ID NO: 13). The
alignment was calculated using the program PILEUP. The location
of the gaps was refined by hand. Capitalized amino acid residues
in the alignment at a given position belong to the amino acid
30 coalition that establish the consensus residue. In bold, beneath
the calculated consensus sequence, the amino acid sequence of

the finally constructed consensus phytase (Fcp) is shown (SEQ ID NO: 14). The gaps in the calculated consensus sequence were filled by hand according to principals stated in Example 1.

5 Figure 2: DNA sequence (SEQ ID NO: 15) of the consensus phytase-1 gene (fcp) and of the primers used for the gene construction. The calculated amino acid sequence (Figure 1, SEQ ID NO: 14) was converted into a DNA sequence using the program BACKTRANSLATE [Devereux, J., Haeberli, P. & Smithies, O. (1984)
10 A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12, 387-395], and the codon frequency table of highly expressed yeast genes (GCG program package, 9.0). The signal peptide of the phytase from *A. terreus* cbs 116.46 was fused to the N-terminus. The amino acid sequence shown in Fig. 2
15 is SEQ ID NO: 16. The bold bases represent the sequences of the oligonucleotides used to generate the gene. The names of the respective oligonucleotides are alternately noted above or below the sequence. The underlined bases represent the start and stop codon of the gene. The bases written in italics represent the
20 two introduced Eco RI sites.

Figure 3: Alignment and consensus sequence of five Basidiomycete phytases. The letters represent the amino acid residues in the one-letter code. The amino acid sequences of the
25 phytases from *Paxillus involutus*, phyA1 (from aa 21; SEQ ID NO: 17; and phyA2 (from aa 21, WO 98/28409; SEQ ID NO: 18); *Trametes pubescens* (from aa 24, WO 98/28409; SEQ ID NO: 19); *Agrocybe pediades* (from aa 19, WO 98/28409; SEQ ID NO: 20); and *Peniophora lycii* (from aa 21, WO 98/28409; SEQ ID NO: 21),
30 starting with the amino acid residues mentioned in parentheses, were used for the alignment and the calculation of the

DDB2.FD "5923450

corresponding consensus sequence called "Basidio" (Example 2; SEQ ID NO: 22). The alignment was performed with the program PILEPUP. The location of the gaps was refined by hand. The consensus sequence was calculated by the program PRETTY. While a
 5 vote weight of 0.5 was assigned to the two *P. involutus* phytases, all other genes were used with a vote weight of 1.0 for the consensus sequence calculation. At positions where the program was not able to determine a consensus residue, the Basidio sequence contains a dash. Capitalized amino acid
 10 residues in the alignment at a given position represent the amino acid coalition that established the consensus residue.

Figure 4: Design of consensus phytase-10 amino acid sequence. By adding the sequence of *Thermomyces lanuginosus*
 15 phytase [Berka, R. M., Rey, M. W., Brown, K. M., Byun, T. & Klotz, A. V. (1998) Molecular characterization and expression of a phytase gene from the thermophilic fungus *Thermomyces lanuginosus*. Appl. Environ. Microbiol. 64, 4423-4427; SEQ ID NO: 23] and the consensus sequence of the phytases from five
 20 Basidiomycetes (SEQ ID NO: 22) to the alignment of Figure 1, an improved consensus sequence was calculated by the program PRETTY. Additionally, the amino acid sequence of *A. niger* T213 was omitted, and a vote weight of 0.5 was assigned to the remaining two *A. niger* phytase sequences. For further
 25 information see Example 2.

Figure 5: DNA and amino acid sequence of consensus phytase-10 (SEQ ID NO: 25, and SEQ ID NO: 26, respectively). The amino acid sequence is written above the corresponding DNA
 30 sequence using the one-letter code. The sequence of the oligonucleotides that were used to assemble the gene are in

000210"59288450

bold letters. The names of the respective oligonucleotides and the amino acids that differ relative to consensus phytase-1 are underlined. The fcp10 gene was assembled from the following oligonucleotides: CP-1, CP-2, CP-3.10, CP-4.10, CP-5.10, CP-6, CP-7.10, CP-8.10, CP-9.10, CP-10.10, CP-11.10, CP-12.10, CP-13.10, CP-14.10, CP-15.10, CP-16.10, CP-17.10, CP18.10, CP-19.10, CP-20.10, CP-21.10, and CP-22.10. The newly synthesized oligonucleotides are additionally marked by the number 10. The phytase contains the following 32 exchanges relative to consensus phytase-1: Y54F, E58A, D69K, D70G, A94K, N134Q, I158V, S187A, Q188N, D197N, S204A, T214L, D220E, L234V, A238P, D246H, T251N, Y259N, E267D, E277Q, A283D, R291I, A320V, R329H, S364T, I366V, A379K, S396A, G404A, Q415E, A437G, A463E. The underlined mutations revealed a stabilizing effect on consensus phytase-1 when tested as single mutations in consensus phytase-1.

Figure 6: Alignment for the design of consensus phytase-11 (SEQ ID NO: 27). In contrast to the design of consensus phytase-10, for the design of the amino acid sequence of consensus phytase-11, all Basidiomycete phytases were used as independent sequences using an assigned vote weight of 0.2 for each Basidiomycete sequence. Additionally, the amino acid sequence of *A. niger* T213 was again used in this alignment.

Figure 7: DNA and amino acid sequence of consensus phytase-1-thermo[8]-Q50T-K91A (SEQ ID NO: 28, and SEQ ID NO: 29, respectively). The amino acid sequence is written above the corresponding DNA sequence using the one-letter code. The replaced amino acid residues (relative to consensus phytase-1) are underlined. The stop codon of the gene is marked by a star (*).

Figure 8: DNA and amino acid sequence of consensus phytase-10-thermo[3]-Q50T-K91A (SEQ ID NO: 30, and SEQ ID NO: 31, respectively). The amino acid sequence is written above the corresponding DNA sequence using the one-letter code. The replaced amino acid residues (relative to consensus phytase-10) are underlined. The stop codon of the gene is marked by a star (*).

Figure 9: DNA and amino acid sequence of *A. fumigatus* ATCC 13073 phytase alpha-mutant Q51T (SEQ ID NO: 32, and SEQ ID NO: 33, respectively). The amino acid sequence is written above the corresponding DNA sequence using the one-letter code. The replaced amino acid residues (relative to *A. fumigatus* ATCC 13073 phytase) are underlined. The stop codon of the gene is marked by a star (*).

Figure 10: DNA and amino acid sequence of consensus phytase-7 (SEQ ID NO: 34, and SEQ ID NO: 35, respectively). The amino acids are written above the corresponding DNA sequence using the one-letter code. The sequence of the oligonucleotides used to assemble the gene are in bold letters. Oligonucleotides and amino acids that were exchanged (relative to consensus phytase-1) are underlined and the corresponding triplets are written in small case letters. The fcp7 gene was assembled from the following oligonucleotides: CP-1, CP-2, CP-3, CP-4.7, CP-5.7, CP-6, CP-7, CP-8.7, CP-9, CP-10.7, CP-11.7, CP-12.7, CP-13.7, CP-14.7, CP-15.7, CP-16, CP-17.7, CP-18.7, CP-19.7, CP-20, CP-21, and CP-22. The newly synthesized oligonucleotides are additionally marked by the number 7. Consensus phytase-7 contains the following 24 exchanges in comparison to the

original consensus phytase-1: S89D, S92G, A94K, D164S, P201S, G203A, G205S, H212P, G224A, D226T, E255T, D256E, V258T, P265S, Q292H, G300K, Y305H, A314T, S364G, M365I, A397S, S398A, G404A, and A405S.

5

Figure 11: Differential scanning calorimetry (DSC) of consensus phytase-1 and consensus phytase-10. The protein samples were concentrated to about 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10°C/min was applied up to 95°C. DSC of consensus phytase-10 (upper graph) yielded a melting temperature of 85.4°C, which is 7.3°C higher than the melting point of consensus phytase-1 (78.1°C, lower graph).

Figure 12: Differential scanning calorimetry (DSC) of consensus phytase-10-thermo[3]-Q50T and consensus phytase-10-thermo[3]-Q50T-K91A. The protein samples were concentrated to ca. 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10°C/min was applied up to 95°C. DSC of consensus phytase-10-thermo[3]-Q50T (upper graph) yielded a melting temperature of 88.6°C, while the melting temperature of consensus phytase-10-thermo-Q50T-K91A was determined to be 89.3°C.

Figure 13: Comparison of the temperature optimum between consensus phytase-1, consensus phytase-10 and consensus phytase-10-thermo[3]-Q50T. For the determination of the temperature optimum, the phytase standard assay of Example 9 was performed at a series of temperatures between 37 and 86°C. The diluted supernatant of transformed *S. cerevisiae* strains was used for the determination. The other components of the supernatant had

no influence on the determination of the temperature optimum: \wedge , consensus phytase-1; \diamond , consensus phytase-10; \blacksquare , consensus phytase 10-thermo[3]-Q50T.

5 Figure 14: pH-dependent activity profile and substrate specificity of consensus phytase-10 and its variants thermo[3]-Q50T and thermo[3]-Q50T-K91A. The phytase activity was determined using the standard assay in appropriate buffers (see Example 9) at different pH-values. Graph a) shows the pH-
10 dependent activity profile of consensus phytase-10 (\square), consensus phytase-10-thermo[3]-Q50T (\bullet), and consensus phytase-10-thermo[3]-Q50T-K91A (\wedge). Graph b) shows the corresponding substrate specificity tested by replacement of phytate in the standard assay by the indicated compounds; open bars, consensus
15 phytase-10; grey bars, consensus phytase-10-thermo[3]-Q50T; dark bars, consensus phytase-10-thermo[3]-Q50T-K91A). The numbers correspond to the following substrates: 1, phytate; 2, p-nitrophenyl phosphate; 3, phenyl phosphate; 4, fructose-1,6-bisphosphate; 5, fructose-6-phosphate; 6, glucose-6-phosphate;
20 7, ribose-5-phosphate; 8, DL-glycerol-3-phosphate; 9, glycerol-2-phosphate; 10, 3-phosphoglycerate; 11, phosphoenolpyruvate; 12, AMP; 13, ADP; 14, ATP.

Figure 15: pH-dependent activity profile and substrate
25 specificity of consensus phytase-1-thermo[8]-Q50T and of consensus phytase-1-thermo[8]-Q50T-K91A. The phytase activity was determined using the standard assay in appropriate buffers (see Example 9) at different pH-values. Graph a) shows the pH-dependent activity profile of the Q50T- (\blacksquare) and the Q50T-K91A-

variant (.). Graph b) shows the corresponding substrate specificities tested by replacement of phytate in the standard assay by the indicated compounds (open bars, consensus phytase-1-thermo[8]-Q50T; filled bars, consensus phytase-1-thermo[8]-Q50T-K91A). The substrates are listed in the legend of Figure 14.

Figure 16: Differential scanning calorimetry (DSC) of consensus phytase-1-thermo[8]-Q50T and consensus phytase-1-thermo[8]-Q50T-K91A. The protein samples were concentrated to ca. 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10°C/min was applied up to 95°C. DSC of consensus phytase-1-thermo[8]-Q50T (upper graph) showed a melting temperature of 84.7°C, while the melting point of consensus phytase-1-thermo[8]-Q50T-K91A was found at 85.7°C.

Figure 17: Comparison of the temperature optimum between consensus phytase-1, consensus phytase-1-thermo[3] and consensus phytase-1-thermo[8]. For the determination of the temperature optimum, the phytase standard assay was performed at a series of temperatures between 37 and 86°C. Protein purified from the supernatant of transformed *S. cerevisiae* strains was used for the determination. ○, consensus phytase-1; □, consensus phytase-1-thermo[3]; ▲, consensus phytase 1-thermo[8].

Figure 18: Comparison of the pH-dependent activity profile and substrate specificity between consensus phytase-1, consensus phytase-7, and the phytase from *A. niger* NRRL 3135. The phytase activity was determined using the standard assay in appropriate buffers (see Example 9) at different pH-values. Graph a) shows

the pH-dependent activity profile of consensus phytase-1 (■), the phytase from *A. niger* NRRL 3135 (○), and of consensus phytase-7 (▲). Graph b) shows the corresponding substrate specificities tested by replacement of phytate in the standard assay by the indicated compounds (black bars, *A. niger* NRRL 3135 phytase; open bars, consensus phytase-1; dashed bars, consensus phytase-7). The substrates are listed in the legend of Figure 14.

Figure 19: Differential scanning calorimetry (DSC) of the phytase from *A. fumigatus* ATCC 13073 and of its stabilized alpha-mutant, which contains the following amino acid exchanges: F55Y, V100I, F114Y, A243L, S265P, and N294D.

The protein samples were concentrated to ca. 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10°C/min was applied up to 95°C. DSC of *A. fumigatus* 13073 phytase (lower graph) revealed a melting temperature of 62.5°C, while the melting point of the alpha-mutant was found at 67.0°C.

Figure 20: Comparison of the temperature optima of *A. fumigatus* 13073 wild-type phytase, its alpha-mutant, and a further stabilized alpha-mutant (E59A-S154N-R329H-S364T-G404A). For the determination of the temperature optimum, the phytase standard assay was performed at a series of temperatures between 37 and 75°C. The diluted supernatant of transformed *S. cerevisiae* strains was used for the determination. The other components of the supernatant had no influence on the determination of the temperature optimum. ○, *A. fumigatus* ATCC 13073 phytase; ▲, *A. fumigatus* ATCC 13073 alpha-mutant; □, *A.*

fumigatus ATCC 13073 alpha-mutant-(E59A-S154N-R329H-S364T-G404A)-Q27T; ■, A. fumigatus ATCC 13073 alpha-mutant-(E59A-S154N-R329H-S364T-G404A)-Q51T-K92A. Q51T and K92A correspond to consensus phytase-1 substitutions Q50T and K91A, respectively.

5

Figure 21: Amino acid sequence of consensus phytase-12 (consphyl2; SEQ ID NO: 36) which contains a number of active site residues transferred from the "basidio" consensus sequence to consensus phytase-10-thermo[3]-Q50T-K91A (underlined).

10

Figure 22: DNA and amino acid sequence of consensus phytase-3-thermo[11]-Q50T. The amino acids are written below the corresponding DNA sequence using the one-letter code.

15

Figure 23: DNA and amino acid sequence of consensus phytase-3-thermo[11]-Q50T-K91A. The amino acids are written below the corresponding DNA sequence using the one-letter code.

Figure 24: DNA and amino acid sequence of consensus phytase-10-thermo[5]-Q50T. The amino acids are written below the corresponding DNA sequence using the one-letter code.

Figure 25: DNA and amino acid sequence of consensus phytase-10-thermo[5]-Q50T-K91A. The amino acids are written below the corresponding DNA sequence using the one-letter code.

The phytase-producing microorganism strains mentioned herein, viz. Paxillus involutus CBS 100231; Peniophora lycii CBS 686.96; Agrocybe pediades CBS 900.96; and Trametes pubescens CBS 100232; were isolated from natural samples originating from, respectively, Denmark; Denmark; Denmark; and Sweden (the Uppsala

000210" 59232460

collection. The samples were collected in November 1992; October 1993; June 1995; and in November 1995, respectively.

Example 1

5 Consensus phytase-1

The amino acid sequence of consensus phytase-1 (fungal consensus phytase, fcp) was designed and calculated as described in Examples 1 and 2 of EP 897985. Table 1 below shows the origin and vote weight of the phytase amino acid sequences used for the
10 design of consensus phytase-1. The consensus phytase-1 sequence was furthermore converted into a DNA sequence as described in Example 3 of EP 897985, and the consensus phytase-1 gene was constructed and cloned as described in Example 4 of EP 897985.

15 Table 1

Origin and vote weight of the phytase amino acid sequences

- phyA from *Aspergillus terreus* 9A-1, aa 27, vote weight 0.5 (Mitchell et al., 1997)
- phyA from *Aspergillus terreus* cbs116.46, aa 27, vote weight
20 0.5 (EP 897985)
- phyA from *Aspergillus niger* var. *awamori*, aa 27, vote weight 0.33 [Piddington, C. S., Houston, C. S., Paloheimo, M., Cantrell, M., Miettinen-Oinonen, A., Nevalainen, H., & Rambosek, J. (1993) The cloning and sequencing of the genes encoding
25 phytase (phy) and pH 2.5-optimum acid phosphatase (aph) from *Aspergillus niger* var. *awamori*. Gene 133, 55-62].
- phyA from *Aspergillus niger* T213 (EP 897985), aa 27, vote weight 0.33
- 30 - phyA from *Aspergillus niger* strain NRRL3135, aa 27, vote weight 0.33 (van Hartingsveldt et al., 1993)

Downloaded from www.aspen.com

- phyA from *Aspergillus fumigatus* ATCC 13073, aa 26, vote weight 0.2 (Pasamontes et al., 1997)
- phyA from *Aspergillus fumigatus* ATCC 32722, aa 26, vote weight 0.2 (EP 897985)
- 5 - phyA from *Aspergillus fumigatus* ATCC 58128, aa 26, vote weight 0.2 (EP 897985)
- phyA from *Aspergillus fumigatus* ATCC 26906, aa 26, vote weight 0.2 (EP 897985)
- phyA from *Aspergillus fumigatus* ATCC 32239, aa 30, vote weight
- 10 0.2 (EP 897985)
- phyA from *Emmericella nidulans* , aa 25, vote weight 1.0 (Pasamontes et al., 1997a)
- phyA from *Talaromyces thermophilus* ATCC 20186, aa 24, vote weight 1.0 (Pasamontes et al., 1997a)
- 15 - phyA from *Myceliophthora thermophila*, aa 19, vote weight 1.0 (Mitchell et al., 1997)

Example 2

20 Design of an improved consensus phytase (consensus phytase-10) amino acid sequence

The alignments used for the design of consensus phytase-10 were calculated using the program PILEUP from the GCG Sequence Analysis Package Release 9.0 (Devereux et al., 1984) with the

25 standard parameters (gap creation penalty 12, gap extension penalty 4). The location of the gaps was refined using a text editor.

The following sequences were used for the alignment of the Basidiomycete phytases starting with the amino acid (aa)

30 mentioned in Table 2:

Table 2

Origin and vote weight of five Basidiomycete phytases used for the calculation of the corresponding consensus amino acid sequence (basidio)

5

- ```
- phyA1 from Paxillus involutus CBS No. 100231, aa 21, vote
weight 0.5 (WO 98/28409)
- phyA2 from Paxillus involutus CBS No. 100231, aa 21, vote
weight 0.5 (WO 98/28409)
```

- ```

10 - phyA from Trametes pubescens CBS No. 100232, aa 24, vote
    weight 1.0 (WO 98/28409)
    - phyA from Agrocybe pediades CBS No. 900.96, aa 19, vote weight
      1.0 (WO 98/28409)
    - phyA from Peniophora lycii CBS No. 686.96, aa 21, vote weight
15 1.0 (WO 98/28409)

```

The alignment is shown in Figure 3.

In Table 3 the genes that were used for the final
20 alignment are listed. The first amino acid (aa) of the sequence
that is used in the alignment is mentioned behind the organism's
designation.

Table 3

25 Origin and vote weight of the phytase sequences used for the
design of consensus phytase-10

- ```

- phyA from Aspergillus terreus 9A-1, aa 27, vote weight 0.5
(Mitchell et al., 1997)
30 - phyA from Aspergillus terreus cbs116.46, aa 27, vote weight
0.5 (EP 897985)

```

- phyA from *Aspergillus niger* var. *awamori*, aa 27, vote weight 0.5 (Piddington et al., 1993)
- phyA from *Aspergillus niger* strain NRRL3135, aa 27, vote weight 0.5 (van Hartingsveldt et al., 1993)
- 5 - phyA from *Aspergillus fumigatus* ATCC 13073, aa 26, vote weight 0.2 (Pasamontes et al., 1997)
- phyA from *Aspergillus fumigatus* ATCC 32722, aa 26, vote weight 0.2 (EP 897985)
- phyA from *Aspergillus fumigatus* ATCC 58128, aa 26, vote weight 10 0.2 (EP 897985)
- phyA from *Aspergillus fumigatus* ATCC 26906, aa 26, vote weight 0.2 (EP 897985)
- phyA from *Aspergillus fumigatus* ATCC 32239, aa 30, vote weight 0.2 (EP 897985)
- 15 - phyA from *Emmericella nidulans* , aa 25, vote weight 1.0 (Pasamontes et al., 1997a)
- phyA from *Talaromyces thermophilus* ATCC 20186, aa 24, vote weight 1.0 (Pasamontes et al., 1997a)
- phyA from *Myceliophthora thermophila*, aa 19, vote weight 1.0 20 (Mitchell et al., 1997)
- phyA from *Thermomyces lanuginosus*, aa 36, vote weight 1.0 (Berka et al., 1998)
- Consensus sequence of five Basidiomycete phytases, vote weight 1.0 (Basidio, Figure 3)

25

The corresponding alignment is shown in Figure 4.

#### Calculation of the amino acid sequence of consensus phytase-10

30 To improve the alignment, we added the original consensus sequence of five phytases from four different Basidiomycetes

000210"59233450

(called Basidio; still containing the undefined sequence positions; see Figure 3), nearly all phytase sequences used for the calculation of the original consensus phytase sequences and one new phytase sequence from the Ascomycete *Thermomyces* 5 *lanuginosus* to a larger alignment.

We set plurality on 2.0 and threshold on 3. The used vote weights are listed in Table 3. The alignment and the corresponding consensus sequence are presented in Figure 4. The new consensus phytase sequence has 32 different amino acids in 10 comparison to the original consensus phytase-1. Positions for which the program PRETTY was not able to calculate a consensus amino acid residue were filled according to rules mentioned in Example 1. None of the residues suggested by the program was replaced.

15 Furthermore, in another calculation, we included all Basidiomycete phytases as single amino acid sequences but assigning a vote weight of 0.2 in the calculation. The corresponding alignment is shown in Figure 6. The calculated consensus amino acid sequence (consensus phytase-11) has the 20 following differences to the sequence of consensus phytase-10. Letter X means that the program was not able to calculate a consensus amino acid; the amino acid in parenthesis corresponds to the amino acid finally included into consensus phytase-10.

D35X (first letter for consensus phytase-10, last letter 25 for consensus phytase-11), X(K)69K, X(E)100E, A101R, Q134N, X(K)153N, X(H)190H, X(A)204S, X(E)220D, E222T, V227A, X(R)271R, H287A, X(D)288D, X(K)379K, X(I)389I, E390X, X(E)415E, X(A)416A, X(R)446L, E463A. The numbering is as in Fig. 5.

We also checked single amino acid replacements suggested 30 by the improved consensus sequences 10 and 11 on their influence

on the stability of the original consensus phytase-1. The approach is described in example 3.

Conversion of the consensus phytase-10 amino acid sequence

5 into a DNA sequence

The first 26 amino acid residues of *A. terreus* cbs116.46 phytase were used as signal peptide and fused to the N-terminus of consensus phytase-10. The used procedure is further described in Example 1.

10 The resulting sequence of the fcp10 gene is shown in Figure 5.

Construction and cloning of the consensus phytase-10 gene (fcp10)

15 The calculated DNA sequence of fcp10 was divided into oligonucleotides of 85 bp, alternately using the sequence of the sense and the anti-sense strand. Every oligonucleotide overlaps 20 bp with the previous and the following oligonucleotide of the opposite strand. The location of all primers, purchased from  
20 Microsynth, Balgach (Switzerland) and obtained in a PAGE-purified form, is indicated in Figure 5.

PCR-Reactions

In three PCR reactions, the synthesized oligonucleotides  
25 were composed to the entire gene. For the PCR, the High Fidelity Kit from Boehringer Mannheim (Boehringer Mannheim, Mannheim, Germany) and the thermo cycler "The Protokoll™" from AMS Biotechnology (Europe) Ltd. (Lugano, Switzerland) were used. The following oligonucleotides were used in a concentration of 0.2  
30 pMol/ml.

Mix 1.10: CP-1, CP-2, CP-3.10, CP-4.10, CP-5.10, CP-6,  
CP-7.10, CP-8.10, CP-9.10, CP-10.10

Mix 2.10: CP-9.10, CP-11.10, CP-12.10, CP-13.10, CP-  
14.10, CP-15.10, CP-16.10, CP-17.10, CP18.10,  
CP-19.10, CP-20.10, CP-21.10, CP-22.10

The newly synthesized oligonucleotides are marked by the  
number 10. Consensus phytase-10 contains the following 32  
exchanges, which are underlined in Figure 5, in comparison to  
the original consensus phytase-1: Y54F, E58A, D69K, D70G, A94K,  
N134Q, I158V, S187A, Q188N, D197N, S204A, T214L, D220E, L234V,  
A238P, D246H, T251N, Y259N, E267D, E277Q, A283D, R291I, A320V,  
R329H, S364T, I366V, A379K, S396A, G404A, Q415E, A437G, A463E.

Four short PCR primers were used for the assembling of the  
oligonucleotides:

CP-a: *Eco RI*  
5'-TATATGAATTCATGGGCGTGTCGTC-3' (SEQ ID NO: 37)

CP-b:  
5'-TGAAAAGTTCATTGAAGGTTTC-3' (SEQ ID NO: 38)

CP-c.10:  
5'-TCTTCGAAAGCAGTACACAAAC-3' (SEQ ID NO: 39)

CP-e: *Eco RI*  
5'-TATATGAATTCCTTAAGCGAAAC-3' (SEQ ID NO: 40)

PCR reaction a: 10  $\mu$ l Mix 1.10 (2.0 pmol of each oligonucleotide)

2  $\mu$ l nucleotides (10 mM of each nucleotide)

2  $\mu$ l primer CP-a (10 pmol/ml)

5 2  $\mu$ l primer CP-c.10 (10 pmol/ml)

10,0  $\mu$ l PCR buffer

0.75  $\mu$ l polymerase mixture (2.6 U)

73.25  $\mu$ l H<sub>2</sub>O

10 PCR reaction b: 10  $\mu$ l Mix 2.10 (2.0 pmol of each oligonucleotide)

2  $\mu$ l nucleotides (10 mM each nucleotide)

2  $\mu$ l primer CP-b (10 pmol/ml)

2  $\mu$ l primer CP-e (10 pmol/ml)

15 10,0  $\mu$ l PCR buffer

0.75  $\mu$ l polymerase mixture (2.6 U)

73.25  $\mu$ l H<sub>2</sub>O

Reaction conditions for PCR reactions a and b:

20            step 1            2 min - 45°C  
              step 2            30 sec - 72°C  
              step 3            30 sec - 94°C  
              step 4            30 sec - 52°C  
              step 5            1 min - 72°C

25

Steps 3 to 5 were repeated 40-times.

The PCR products (670 and 905 bp) were purified by agarose gel electrophoresis (0.9% agarose), followed by gel extraction  
 30 (QIAEX II Gel Extraction Kit, Qiagen, Hilden, Germany). The purified DNA fragments were used for the PCR reaction c.

PCR reaction c: 6  $\mu$ l PCR product of reaction a  $\approx$ 50 ng)  
6  $\mu$ l PCR product of reaction b  $\approx$ 50 ng)  
2  $\mu$ l primer CP-a (10 pmol/ml)  
5 2  $\mu$ l primer CP-e (10 pmol/ml)  
10,0  $\mu$ l PCR buffer  
0.75  $\mu$ l polymerase mixture (2.6 U)  
73.25  $\mu$ l H<sub>2</sub>O

10 Reaction conditions for PCR reaction c:

step 1 2 min - 94°C  
step 2 30 sec - 94°C  
step 3 30 sec - 55°C  
step 4 1 min - 72°C

15

Steps 2 to 4 were repeated 31-times.

The resulting PCR product (1.4 kb) was purified as mentioned above, digested with EcoRI, and ligated in an EcoRI-  
20 digested and dephosphorylated pBsk(-)-vector (Stratagene, La Jolla, CA, USA). 1  $\mu$ l of the ligation mixture was used to transform E. coli XL-1 competent cells (Stratagene, La Jolla, CA, USA). All standard procedures were carried out as described by Sambrook et al. (1987). The DNA sequence of the constructed  
25 gene (fcp10) was checked by sequencing as known in the art.

Example 3

Increasing the thermostability of consensus phytase-1 by  
introduction of single mutations suggested by the amino  
acid sequences of consensus phytase-10 and consensus  
phytase-11

5

In order to increase the thermostability of homologous genes, it is also possible to test the stability effect of each differing amino acid residue between the protein of interest and  
10 the calculated consensus sequence and to combine all stabilizing mutations into the protein of interest. We used the consensus phytase-1 as protein of interest and tested the effect on the protein stability of 34 amino acid residues that differ relative to consensus phytase-10 and/or -11 by single site-directed  
15 mutagenesis.

To construct muteins for expression in *A. niger*, *S. cerevisiae*, or *H. polymorpha*, the corresponding expression plasmid containing the consensus phytase-1 gene was used as template for site-directed mutagenesis (see Examples 6-8).  
20 Mutations were introduced using the "quick exchange™ site-directed mutagenesis kit" from Stratagene (La Jolla, CA, USA) following the manufacturer's protocol and using the corresponding primers. All mutations made and the corresponding primers are summarized in Table 4. Plasmids harboring the  
25 desired mutation were identified by DNA sequence analysis as known in the art.

DDB2.F0"59232750

Table 4Primers used for site-directed mutagenesis of consensus phytase-  
1

Exchanged bases are highlighted in bold. The introduction of a  
5 restriction site is marked above the sequence. When a  
restriction site is written in parenthesis, the mentioned site  
was destroyed by introduction of the mutation.

| mutation | Primer set                                                                                                                                    |
|----------|-----------------------------------------------------------------------------------------------------------------------------------------------|
|          | <i>Kpn I</i>                                                                                                                                  |
| 10 Q50T  | 5' -CACTTGTGGGG <b>TAC</b> TACTCTCCATACTTCTC-3' (SEQ ID NO: 41)<br>5' -GAGAAGTATGGAGAGT <b>AGG</b> TACCCACAAGT-3'                             |
| Y54F     | 5' -GGTCAATACTCTCCATTCT <b>T</b> CTCTTTGGAAG-3' (SEQ ID NO: 42)                                                                               |
| 15       | 5' -CTTCCAAAGAGA <b>AGA</b> ATGGAGAGTATTGACC-3'                                                                                               |
| E58A     | 5' -CATACTTCTCTTTGG <b>C</b> AGACGAATCTGC-3' (SEQ ID NO: 43)<br>5' -GCAGATTCTCT <b>G</b> CCAAAGAGAAGTATG-3'                                   |
|          | <i>Aat II</i>                                                                                                                                 |
| 20 D69K  | 5' -CTCCAGAC <b>GT</b> CCCA <b>AA</b> GGACTGTAGAGTTAC-3' (SEQ ID NO: 44)<br>5' -GTAAGTCTACAGTC <b>CTT</b> TGG <b>G</b> ACGTCTGGAG-3'          |
|          | <i>Aat II</i>                                                                                                                                 |
| 25 D70G  | 5' -CTCCAGAC <b>GT</b> CC <b>C</b> AGAC <b>G</b> GCTGTAGAGTTAC-3' (SEQ ID NO: 45)<br>5' -GTAAGTCTACAG <b>CC</b> GTCTGG <b>G</b> ACGTCTGGAG-3' |
| K91A     | 5' -GATACCCA <b>ACT</b> TCTTCT <b>G</b> CGTCTAAGGCTTACTCTG-3'<br>(SEQ ID NO: 46)                                                              |
| 30       | 5' -CAGAGTAAGCCTTAGAC <b>G</b> CAGAAGAAGTTGGGTATC-3'                                                                                          |
|          | <i>Sca I</i>                                                                                                                                  |
| A94K     | 5' -CTTCTAAGTCTAAG <b>AA</b> GTACTCTGCTTTG-3' (SEQ ID NO: 47)<br>5' -CAAAGCAGAGT <b>ACT</b> TTCTTAGACTTAGAAG-3'                               |
| 35       |                                                                                                                                               |
| A101R    | 5' -GCTTACTCTGCTTTGATTGAAC <b>CG</b> GATTCAAAAGAACGCTAC-3'<br>(SEQ ID NO: 48)<br>5' -GTAGCGTTCTTTTGAAT <b>CC</b> GTTCAATCAAAGCAGAGTAAGC-3'    |
| 40 N134Q | 5' -CCATTTCGGTGAAC <b>AG</b> C <b>AA</b> ATGGTTAACTC-3' (SEQ ID NO: 49)<br>5' -GAGTTAACCATT <b>TG</b> CT <b>G</b> TTCACCGAATGG-3'             |

00020"5322450

Nru I

K153N 5'-GATACAAGGCT**CTCGCG**GAGAAACATTGTTC -3' (SEQ ID NO: 50)  
5'-GGAACAAT**GTTTCTCGCG**GAGAGCCTTGTATC-3'

## BSS HI

5 *Bss* HI  
I158V 5' -GATTGTTCCATTC**GTGCGCG**CTTCTGGTTC-3' (SEQ ID NO: 51)  
5' -GAACCAGAAG**CGCGC**ACGAATGGAACAATC-3'

*Apa* I

10 S187A 5' -GGCTGACCCAGG**GGGCC**CAACCACACCAAGC-3' (SEQ ID NO: 53)  
5' -GCTTGGTGTGGTT**GGGCC**CCTGGGTCAGCC-3'

Bcl I

D197N 5' -CTCCAGTTATTA**AAC**GTGATCATTCCAGAAGG-3' (SEQ ID NO: 52)  
15 5' -CCTTCTGGAATGATCAC**GTT**AATAACTGGAG-3'

Nco I

T214L 5' -CACTTTGGACCA**TGGTCT**TTTGTACTGCTTTTCG-3' (SEQ ID NO: 54)  
5' -CGAAAGCAGTACAA**AGACCA**TGGTCCAAAGTG-3'

## Avr II

E222T 5' -GCTTTCGAAGACTCT**ACCC**TAGGTGACGACGTTG-3' (SEQ ID NO: 55)  
5' -CAACGTCGTCACCTA**GGG**TAGAGTCTTCGAAAGC-3'

25 V227A 5' -GGTGACGACG**CTGAAGCTAACTTCAC**-3' (SEQ ID NO: 56)  
5' -GTGAAGTTAG**CTTCAGCGTCGTCACC**-3'

Sac II

30 L234V 5' -CTAACTTCAC**CGCGGT**GTTCGCTCCAG-3' (SEQ ID NO: 57)  
5' -CTGGAGCGAACA**CCGCGGT**GAAGTTAG-3'

A238P 5' -GCTTTGTTTCGCTCC**A**CTATTAGAGCTAGATTGG-3' (SEQ ID NO: 58)

35 5' -CCAATCTAGCTCTAATAG**G**TGGAGCGAACAAAGC-3'

Ира I

T251N 5' -GCCAGGTGTTAACTTG**ACT**GACGAAG-3' (SEQ ID NO: 59)  
5' -TTCGTCAGTCAAG**GT**TAAACACCTGGC-3'

## Aat II

Y259N      5' - GACGAAGACGT**TC**GTTAACTTGATGGAC - 3' (SEQ ID NO: 60)  
             5' - GTCCATCAAGTTAACGACGTCTTCGTC - 3'

Asp I

45                      Asp I  
E267D     5' -GTCCATTGCA**C**ACTGTCGCTAGAACTTC-3' (SEQ ID NO: 61)  
          5' -GAAGTTCTAGCGACAGT**G**TCCAATGGAC-3'

E277Q 5' -CTGACGCTACTCAGCTGTCTCCATTC-3' (SEQ ID NO: 62)  
5' -GAATGGAGACAGCTGAGTAGCGTCAG-3'

5 A283D 5' -GTCTCCATTCTGTGATTTGTTCACTCAC-3' (SEQ ID NO: 63)  
5' -GTGAGTGAACAAATCACAGAATGGAGAC-3'

*Ksp I*

10 H287A 5' -GCTTTGTTCA**CCGCG**GACGAATGGAG-3' (SEQ ID NO: 64)  
5' -CTCCATTTCGT**CCGCG**GTGAACAAAGC-3'

*Bam HI*

15 R291I 5' -CACGACGAATGGAT**CCA**TACGACTAC-3' (SEQ ID NO: 65)  
5' -GTAGTCGTATTG**GAT**CCATTCGTCGTG-3'

*Bsi WI*

Q292A 5' -GACGAATGGAGAG**CGT**ACGACTACTTG-3' (SEQ ID NO: 66)  
5' -CAAGTAGTCGT**ACG**CTCTCCATTCGTC-3'

*Hpa I*

20 A320V 5' -GGTGTGGTTTCGT**TAA**CGAATTGATTGC-3' (SEQ ID NO: 67)  
5' -GCAATCAATTCGT**TAA**CGAAACCAACACC-3'

*(Bgl II)*

25 R329H 5' -GCTAGATTGACT**CACT**CTCCAGTTCAAG-3' (SEQ ID NO: 68)  
5' -CTTGAAGTGGAGAG**TGAG**TCAATCTAGC-3'

*Eco RV*

30 S364T 5' -CTCACGACAAC**ACTATGATA**TCTATTTTCTTC-3' (SEQ ID NO: 69)  
5' -GAAGAAAATAG**ATATCATAGT**GTTGTCGTGAG-3'

*Nco I*

35 I366V 5' -CGACAACTC**CATGG**TTTCTATTTTCTTCGC-3' (SEQ ID NO: 70)  
5' -GCGAAGAAAATAGAAAC**CATGG**AGTTGTCG-3'

*Kpn I*

A379K 5' -GTACAACGGTAC**CAAG**CCATTGTCTAC-3' (SEQ ID NO: 71)  
5' -GTAGACAATGG**CTTGG**TACCGTTGTAC-3'

40 S396A 5' -CTGACGGTTAC**GCTG**CTTCTTGGAC-3' (SEQ ID NO: 72)  
5' -GTCCAAGAAGCAG**CGTA**ACCGTCAG-3'

G404A 5' -CTGTTCCATT**CGTG**CTAGAGCTTAC-3' (SEQ ID NO: 73)  
5' -GTAAGCTCTAGCAG**CGA**ATGGAACAG-3'

45 Q415E 5' -GATGCAATGT**GAAG**CTGAAAAGGAACC-3' (SEQ ID NO: 74)  
5' -GGTTCCTTTT**CAGCTT**CACATTGCATC-3'

00020" 59232450

Sal I

A437G     5'-CACGGTTGTGGTGT**CG**ACAAGTTGGG-3' (SEQ ID NO: 75)  
          5'-CCCAACTTGT**CG**ACACCACAACCGTG-3'

5

Mun I

A463E     5'-GATCTGGTGG**CAATTG**GG**AG**GAATGTTTCG-3' (SEQ ID NO: 76)  
          5'-CGAAACATT**CTCCCAATTG**CCACCAGATC-3'

10 and, accordingly, for other mutations.

          The temperature optimum of the purified phytases,  
expressed in *Saccharomyces cerevisiae* (Example 7), was  
determined as outlined in Example 9. Table 5 shows the effect of  
15 each mutation introduced on the stability of consensus phytase-  
1.

Table 5

Stability effect of the individual amino acid replacements in  
20 consensus phytase-1

+ or - means a positive, respectively, negative effect on the  
protein stability up to 1°C, ++ and -- means a positive,  
respectively, negative effect on the protein stability between 1  
and 3°C; the numbers 10 or 11 in parentheses indicate the  
25 consensus phytase sequence that suggested the amino acid  
replacement.

000210-5333-60

| stabilizing |        | neutral    |        | destabilizing |        |
|-------------|--------|------------|--------|---------------|--------|
| mutation    | effect | mutation   | effect | mutation      | effect |
| E58A (10)   | +      | D69A       | ±      | Y54F (10)     | -      |
| D69K (11)   | +      | D70G (10)  | ±      | V73I          | -      |
| D197N (10)  | +      | N134Q (10) | ±      | A94K (10)     | -      |
| T214L (10)  | + +    | G186H      | ±      | A101R (11)    | -      |
| E222T (11)  | + +    | S187A (10) | ±      | K153N (11)    | -      |
| E267D (10)  | +      | T214V      | ±      | I158V (10)    | - -    |
| R291I       | +      | T251N (10) | ±      | G203A         | - -    |
| R329H (10)  | +      | Y259N (10) | ±      | G205S         | -      |
| S364T (10)  | + +    | A283D (10) | ±      | A217V         | -      |
| A379K (11)  | +      | A320V (10) | ±      | V227A (11)    | - -    |
| G404A (10)  | + +    | K445T      | ±      | L234V (10)    | -      |
|             |        | A463E (10) | ±      | A238P (10)    | - -    |
|             |        |            |        | E277Q (10)    | -      |
|             |        |            |        | H287A (11)    | -      |
|             |        |            |        | Q292A         | -      |
|             |        |            |        | I366V (10)    | -      |
|             |        |            |        | S396A (10)    | - -    |
|             |        |            |        | Q415E (11)    | -      |
|             |        |            |        | A437G (10)    | - -    |
|             |        |            |        | E451R         | - -    |

We combined eight positive mutations (E58A, D197N, E267D, R291I, R329H, S364T, A379K, G404A) in consensus phytase-1 thermo[8], using the primers and the technique mentioned above in this example. Furthermore, the mutations Q50T and/or K91A were introduced which mainly influence the catalytic characteristics of phytase (see patent applications EP 897010 and EP 897985, as well as Example 9). The DNA and amino acid sequence of the resulting phytase (consensus phytase-1-thermo[8]-Q50T-K91A) are shown in Figure 7. In this way, the temperature optimum and the melting point of the consensus phytase were increased by 7°C (Figures 15, 16, 17).

In a further consensus protein, we combined eleven positive mutations (E58A, D69K, D197N, T214L, E222T, E267D, R291I, R329H, S364T, A379K, G404A) in consensus phytase-1 thermo[11]. Furthermore, the mutations Q50T and/or K91A were

introduced. In this way, the melting temperature was increased by another 3-4°C when compared to consensus phytase-1 thermo[8].

Using the results of Table 5, we further improved the thermostability of consensus phytase-10 by the back mutations K94A, V158I, and A396S, the reverse of which (A94K, I158V, and S396A) revealed a strong negative influence on the stability of consensus phytase-1. The resulting protein was called consensus phytase-10-thermo[3]. SEQ ID NO: 26 plus the three mutations K94A, V158I, and A396S. Furthermore, we introduced the mutations Q50T and K91A that mainly influence the catalytic characteristics of consensus phytase (see patent applications EP 897010 and EP 897985, as well as Example 9 and Figures 14 and 15). The resulting DNA and amino acid sequence are shown in Figure 8. The optimized phytase showed a 4°C higher temperature optimum and melting point than consensus phytase-10 (Figures 12 and 13). Furthermore, the phytase has also a strongly increased specific activity with phytate as substrate of 250 U/mg at pH 5.5 (Figure 14).

In a still further consensus protein, two additional mutations were introduced into consensus phytase-10 thermo[3] (E222T, G437A) which yielded consensus phytase-10 thermo[5]. Furthermore, the mutations Q50T and/or K91A were introduced. In this way, the melting temperature was increased by another 1-2°C when compared to consensus phytase-10 thermo[3].

Example 4

Stabilization of the phytase of *A. fumigatus* ATCC 13073 by replacement of amino acid residues with the corresponding consensus phytase-1 and/or consensus phytase-10 residues

5

At six amino acid sequence positions where *A. fumigatus* 13073 phytase is the only or nearly the only phytase in the alignment of Figure 1 that does not contain the corresponding consensus phytase amino acid residue, the non-consensus amino acid residue was replaced by the consensus one. The following amino acids were substituted in *A. fumigatus* 13073 phytase, containing additionally the Q51(24)T substitution (influencing the catalytic properties and corresponding to the Q50T substitution in the consensus phytases) and the signal sequence of *A. terreus* cbs116.46 phytase (see European Patent Application No. 0897010, and Figure 9): F55(28)Y, V100(73)I, F114(87)Y, A243(220)L, S265(242)P, N294(282)D. The numbers in parentheses refer to the numbering in Figure 1.

In a second round, four of the seven stabilizing amino acid exchanges (E58A, R329H, S364T, G404A) identified in consensus phytase-10 and tested as single mutations in consensus phytase-1 (Table 5) were additionally introduced into the *A. fumigatus* alpha-mutant. Furthermore, the amino acid replacement S154N, shown to reduce the protease susceptibility of the phytase, was introduced.

The mutations were introduced as described in Example 3 (see Table 6) and expressed as described in Examples 6 to 8. The resulting *A. fumigatus* 13073 phytase variants were called alpha-mutant (i.e. the *A. fumigatus* ATCC 13073 phytase with the substitutions Q24T, F28Y, V73I, F87Y, A220L, S242P, N282D) and "optimized" alpha-mutant (i.e. the *A. fumigatus* alpha-mutant

000210"59283450

having the additional substitutions E59A-S154N-R329H-S364T-G404A). K92A is an additional preferred mutation.

The temperature optimum (60°C, Figure 20) and the melting temperature (67.0°C, Figure 19) of the *A. fumigatus* 13073 alpha-5 mutant phytase were increased by 5-7°C in comparison to the values of the wild-type phytase (temperature optimum: 55°C, T<sub>m</sub>: 60°C). The five additional amino acid replacements further increased the temperature optimum by 3°C (Figure 20).

# 10 Table 6

## Mutagenesis primers for the stabilization of *A. fumigatus* ATCC 13073 phytase

|    | Mutation | Primer                                                                                                               |
|----|----------|----------------------------------------------------------------------------------------------------------------------|
| 15 | F55Y     | 5'-CACGTACTCGCCATA <b>ACT</b> TTTTCGCTCGAG-3' (SEQ ID NO: 77)<br>5'-CTCGAGCGAAAAGT <b>AT</b> GGCGAGTACGTG-3'         |
|    |          | ( <i>Xho</i> I)                                                                                                      |
| 20 | E58A     | 5'-CCATACTTTTTCGCTCG <b>CG</b> GACGAGCTGTCCGTG-3' (SEQ ID NO: 78)<br>5'-CACGGACAGCTCGTCC <b>CG</b> GAGCGAAAAGTAGG-3' |
|    | V100I    | 5'-GTATAAGAAGCTT <b>ATT</b> ACGGCGATCCAGGCC-3' (SEQ ID NO: 79)<br>5'-GGCCTGGATCGCCGTA <b>ATA</b> AGCTTCTTATAC-3'     |
| 25 | F114Y    | 5'-CTTCAAGGGCAAGT <b>TAC</b> GCCTTTTGAAGACG-3' (SEQ ID NO: 80)<br>5'-CGTCTTCAAAAAGG <b>CGT</b> ACTTGCCCTTGAAG-3'     |
| 30 | A243L    | 5'-CATCCGAGCTCGC <b>CTC</b> GAGAAGCATCTTC-3' (SEQ ID NO: 81)<br>5'-GAAGATGCTTCTC <b>GAG</b> GCGAGCTCGGATG-3'         |
| 35 | S265P    | 5'-CTAATGGA TGTGT <b>CCG</b> TTTGATACGGTAG-3' (SEQ ID NO: 82)<br>5'-CTACCGTATCAA <b>ACG</b> GACACATGTCCATTAG-3'      |
|    | N294D    | 5'-GTGGAAGAAGT <b>ACG</b> ACTACCTTCAGTC-3' (SEQ ID NO: 83)<br>5'-GACTGAAGGTAGT <b>TCG</b> TACTTCTTCCAC-3'            |

(Mlu I)

R329H 5'-GCCCGGTTGACGCAATTCGCCAGTGCAGG-3' (SEQ ID NO: 84)  
 5'-CCTGCACTGGCGAATGCGTCAACCGGGC-3'

Nco I

5 S364T 5'-CACACGACAACACCATGGTTTCCATCTTC-3' (SEQ ID NO: 85)  
 5'-GAAGATGGAAACCATGGTGTGTGTCGTGTG-3'

(Bss HI)

10 G404A 5'-GTGGTGCCTTTTCGCCGCGAGCCTACTTC-3' (SEQ ID NO: 86)  
 5'-GAAGTAGGCTCGCGCGCGAAAGGCACCAC-3'

Example 5

Introduction of the active site amino acid residues of A.  
niger NRRL 3135 phytase into consensus phytase-1

15 We used the crystal structure of *Aspergillus niger* NRRL 3135 phytase to define all active site amino acid residues (see Example 1, and EP 897010). Using the alignment of Figure 1, we replaced the following active site residues and additionally  
 20 the non-identical adjacent ones of consensus phytase-1 by those of A. niger phytase:

S89D, S92G, A94K, D164S, P201S, G203A, G205S, H212P, G224A, D226T, E255T, D256E, V258T, P265S, Q292H, G300K, Y305H, A314T, S364G, M365I, A397S, S398A, G404A, and A405S.

25 The new consensus phytase-7 protein sequence was backtranslated into a DNA sequence (Figure 10) as described in Example 1. The corresponding gene (fcp7) was generated as described in Example 1 using the following oligonucleotide mixes:

30 Mix 1.7: CP-1, CP-2, CP-3, CP-4.7, CP-5.7, CP-6, CP-7, CP-8.7, CP-9, CP-10.7

Mix 2.7: CP-9, CP-10.7, CP-11.7, CP-12.7, CP-13.7, CP-14.7, CP-15.7, CP-16, CP-17.7, CP-18.7, CP-19.7, CP-20, CP-21, CP-22.

The DNA sequences of the oligonucleotides are indicated in Figure 10. The newly synthesized oligonucleotides are additionally marked by the number 7. After assembling of the 5 oligonucleotides using the same PCR primers as mentioned in Example 1, the gene was cloned into an expression vector as described in Examples 6-8.

The pH-profile of the enzyme determined after expression in *H. polymorpha* and purification was very similar to that of *A. niger* phytase (see Figure 18).

### Example 6

Expression of the consensus phytase genes in Hansenula polymorpha

15 The phytase expression vectors used to transform *H. polymorpha* RB11 [Gellissen, G., Hollenberg, C. P., Janowicz, Z. A. (1994) Gene expression in methylotrophic yeasts, in Smith, A. (ed.) Gene expression in recombinant microorganisms. Dekker, New York, pp. 395-439] were constructed by inserting the Eco RI fragment  
20 of pBsk-fcp or variants thereof into the multiple cloning site of the *H. polymorpha* expression vector pFPMT121, which is based on an *ura3* selection marker from *S. cerevisiae*, a formate dehydrogenase (FMD) promoter element and a methanol oxidase (MO) terminator element from *H. polymorpha*. The 5' end of the fcp  
25 gene is fused to the FMD promoter, the 3' end to the MOX terminator (Gellissen et al., Appl. Microbiol. Biotechnol. 46, 46-54, 1996; EP 299108). The resulting expression vectors are designated pFPMTfcp, pFPMTfcp10, and pFPMTfcp7.

The constructed plasmids were propagated in E. coli. 30 Plasmid DNA was purified using standard state of the art procedures. The expression plasmids were transformed into the H.

polymorpha strain RB11 deficient in orotidine-5'-phosphate decarboxylase (ura3) using the procedure for preparation of competent cells and for transformation of yeast as described in Gellissen et al. (1996). Each transformation mixture was plated  
5 on YNB medium (0.14% w/v Difco YNB and 0.5% ammonium sulfate) containing 2% glucose and 1.8% agar, and incubated at 37 °C. After 4 to 5 days individual transformant colonies were picked and grown in the liquid medium described above for 2 days at 37 °C. Subsequently, an aliquot of this culture was used to  
10 inoculate fresh vials with YNB-medium containing 2% glucose. After seven further passages in selective medium, the expression vector had integrated into the yeast genome in multimeric form. Subsequently, mitotically stable transformants were obtained by two additional cultivation steps in 3 ml non-selective liquid  
15 medium (YPD, 2% glucose, 10 g/l yeast extract, and 20 g/l peptone). In order to obtain genetically homogeneous recombinant strains, an aliquot from the last stabilization culture was plated on a selective plate. Single colonies were isolated for analysis of phytase expression in YNB containing 2% glycerol  
20 instead of glucose to derepress the FMD promoter. Purification of the consensus phytases was done as described in Example 7.

#### Example 7

Expression of the consensus phytase genes in Saccharomyces  
25 cerevisiae and purification of the phytases from the  
culture supernatant

The consensus phytase genes were isolated from the corresponding Bluescript-plasmid (pBsk-fcp, pBSK-fcp10, pBsk-  
30 fcp7) and ligated into the Eco RI sites of the expression cassette of the Saccharomyces cerevisiae expression vector pYES2

(Invitrogen, San Diego, CA, USA) or subcloned between the shortened GAPFL (glyceraldehyde-3-phosphate dehydrogenase) promoter and the pho5 terminator as described by Janes et al., Curr. Genet. 18, 97-103. The correct orientation of the gene was checked by PCR. Transformation of *S. cerevisiae* strains, e.g. INVSc1 (Invitrogen, San Diego, CA, USA), was done according to Hinnen et al., Proc. Natl. Acad. Sci. USA 75, 1929-1933 (1978). Single colonies harboring the phytase gene under the control of the GAPFL promoter were picked and cultivated in 5 ml selection medium [SD-uracil; Sherman, J. P., Finck, G. R. & Hicks, J. B. (1986) Laboratory course manual for methods in yeast genetics. Cold Spring Harbor University] at 30°C under vigorous shaking (250 rpm) for one day. The preculture was then added to 500 ml YPD medium (Sherman et al., 1986) and grown under the same conditions. Induction of the gall promoter was done according to the manufacturer's instructions. After four days of incubation, the cell broth was centrifuged (7000 rpm, GS3 rotor, 15 min, 5°C) to remove the cells, and the supernatant was concentrated by way of ultrafiltration in Amicon 8400 cells (PM30 membranes; Grace AG, Wallizeller, Switzerland) and ultrafree-15 centrifugal filter devices (Biomax-30K, Millipore, Bedford, MA, USA). The concentrate (10 ml) was desalted on a 40 ml Sephadex G25 Superfine column (Pharmacia Biotech, Freiburg, Germany), with 10 mM sodium acetate, pH 5.0, serving as elution buffer. The desalted sample was brought to 2 M  $(\text{NH}_4)_2\text{SO}_4$  and directly loaded onto a 1 ml Butyl Sepharose 4 Fast Flow hydrophobic interaction chromatography column (Pharmacia Biotech, Feiburg, Germany) which was eluted with a linear gradient from 2 M to 0 M  $(\text{NH}_4)_2\text{SO}_4$  in 10 mM sodium acetate, pH 5.0. Phytase was eluted in the breakthrough, concentrated and loaded on a 120 ml Sephacryl S-300 gel permeation chromatography column (Pharmacia Biotech,

Freiburg, Germany). Consensus phytases -1, -7 and -10 eluted as a homogeneous symmetrical peak and were shown by SDS-PAGE to be approx. 95% pure.

5        Example 8

Expression of the consensus phytase genes in Aspergillus niger

The Bluescript-plasmids pBsk-fcp, pBsk-fcp10, and pBsk-fcp7 were used as template for the introduction of a Bsp HI-site upstream of the start codon of the genes and an Eco RV-site downstream of the stop codon. The Expand<sup>TM</sup> High Fidelity PCR Kit (Boehringer Mannheim, Mannheim, Germany) was used with the following primers:

15 Primer Asp-1:

*Bsp HI*

5'-TATATCATGAGCGTGTTCGTCGTGCTACTGTTC-3' (SEQ ID NO: 87)

Primer Asp-2 used for cloning of fcp and fcp7:

20

*Eco RV*

3'-ACCCGACTTACAAAGCGAATTCTATAGATATAT-5' (SEQ ID NO: 88)

Primer Asp-3 used for cloning of fcp10:

25

*Eco RV*

3'-ACCCTTCTTACAAAGCGAATTCTATAGATATAT-5' (SEQ ID NO: 89)

The reaction was performed as described by the supplier.  
30 The PCR-amplified fcp-genes had a new Bsp HI site at the start codon, introduced by primer Asp-1, which resulted in a

replacement of the second amino acid residue glycine by serine. Subsequently, the DNA-fragment was digested with Bsp HI and Eco RV and ligated into the Nco I site downstream of the glucoamylase promoter of *Aspergillus niger* (glaA) and the Eco RV site upstream of the *Aspergillus nidulans* tryptophan C terminator (trpC) (Mullaney et al., 1985). After this cloning step, the genes were sequenced to detect possible errors introduced by PCR. The resulting expression plasmids, which basically correspond to the pGLAC vector as described in Example 9 of EP 684313, contained the orotidine-5'-phosphate decarboxylase gene (pyr4) of *Neurospora crassa* as a selection marker. Transformation of *Aspergillus niger* and expression of the consensus phytase genes was done as described in EP 684313. The consensus phytases were purified as described in Example 7.

#### Example 9

#### Determination of phytase activity and of the pH and temperature optima

This example relates i.a. to the determination of phytase activity and of the temperature optimum. Various phytases have been tested.

The phytase of *Aspergillus niger* NRRL 3135 was prepared as described in EP 420358 and by van Hartingsveldt et al. (Gene 127, 87-94, 1993).

The phytases of *Aspergillus fumigatus* ATCC 13073, *Aspergillus terreus* 9A-1, *Aspergillus terreus* cbs116.46, *Emmericella nidulans*, *Myceliophthora thermophila*, and *Talaromyces thermophilus* were prepared as described in EP-0897985 and in the references therein.

The remaining phytases tested were prepared as described herein.

Consensus phytase-1-thermo(8) designates a variant of consensus phytase-1, which further comprises the eight mutations which are underlined in the legend to Figure 5. Consensus phytase-1 is shown in Fig. 1 (SEQ ID NO: 14) without signal peptide, and in Fig. 2 (SEQ ID NO: 16) with the signal peptide.

Phytase activity was determined basically as described by Mitchell et al. (1997). The activity was measured in an assay mixture containing 0.5% phytic acid ( $\approx 5$  mM) in 200 mM sodium acetate, pH 5.0. After 15 min of incubation at 37°C, the reaction was stopped by addition of an equal volume of 15% trichloroacetic acid. The liberated inorganic phosphate was quantified by mixing 100  $\mu$ l of the assay mixture with 900  $\mu$ l H<sub>2</sub>O and 1 ml of 0.6 M H<sub>2</sub>SO<sub>4</sub>, 2% ascorbic acid and 0.5% ammonium molybdate. Standard solutions of potassium phosphate were used as reference. One unit of enzyme activity was defined as the amount of enzyme that releases 1  $\mu$ mol phosphate per minute at 37°C. The protein concentration was determined using the enzyme extinction coefficient at 280 nm calculated according to Pace et al. [Pace N. C., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. Prot. Sci. 4, 2411-2423]: 1 absorption unit (1 OD) at 280 nm corresponds to 1.101 mg/ml of consensus phytase-1, 1.068 mg/ml of consensus phytase-7, and 1.039 mg/ml of consensus phytase-10.

In case of pH-optimum curves, the purified enzymes were diluted in 10 mM sodium acetate, pH 5.0. Incubations were started by mixing aliquots of the diluted protein with an equal volume of 1% phytic acid ( $\approx 10$  mM) in a series of different buffers: 0.4 M glycine/HCl, pH 2.5; 0.4 M acetate/NaOH, pH 3.0, 3.5, 4.0, 4.5, 5.0, 5.5; 0.4 M imidazole/HCl, pH 6.0, 6.5; 0.4 M Tris/HCl pH 7.0, 7.5, 8.0, 8.5, 9.0. Control experiments showed

that pH was only slightly affected by the mixing step. Incubations were performed for 15 min at 37°C as described above.

For determination of the substrate specificities of the 5 phytases, phytic acid in the assay mixture was replaced by 5 mM concentrations of the respective phosphate compounds. Besides, the activity tests were performed as described above.

For determination of the temperature optimum, enzyme (100  $\mu$ l) and substrate solution (100  $\mu$ l) were pre-incubated for 5 min 10 at the given temperature. The reaction was started by addition of the substrate solution to the enzyme. After 15 min of incubation, the reaction was stopped with trichloroacetic acid, and the amount of phosphate released was determined.

The pH-optimum of consensus phytase-1 was around pH 6.0- 15 6.5 (70 U/mg). Introduction of the Q50T mutation shifted the pH-optimum to pH 6.0 (130 U/mg). Introduction of the K91A mutation further shifted the pH optimum into the more acidic pH-range. Comparable effects of the Q50T and K91A mutations were also observed for consensus phytase-10 and for further 20 stabilized consensus phytase variants (Figures 14 and 15).

Consensus phytase-7, which was constructed to transfer the catalytic characteristics of *A. niger* NRRL 3135 phytase to consensus phytase-1, had a pH-profile very similar to that of *A. niger* NRRL 3135 phytase (see Figure 18). The substrate 25 specificity also resembled more that of *A. niger* NRRL 3135 phytase than that of consensus phytase-1.

The temperature optimum of consensus phytase-1 (71°C) was 16-26°C higher than the temperature optima of the wild-type phytases (45-55°C, Table 7) that were used to calculate the 30 consensus sequence. The improved consensus phytase-10 showed a further increase of its temperature optimum to 80°C (Figure 13).

The temperature optimum of consensus phytase-1-thermo[8] was found to be in the same range (78°C) when using the supernatant of an overproducing *S. cerevisiae* strain. The highest temperature optimum reached of 82°C was determined for consensus phytase-10-thermo[3]-Q50T-K91A. Table 7

Temperature optima and Tm-values of consensus phytase and of the phytases from *A. fumigatus*, *A. niger*, *E. nidulans*, and *M. thermophila*.

10

The determination of the temperature optimum was performed as described in Example 9. The Tm-values were determined by differential scanning calorimetry as described in Example 10.

| Phytase                                  | Optimum temperature<br>(°C) | Tm (°C) |
|------------------------------------------|-----------------------------|---------|
| <i>Aspergillus niger</i><br>NRRL 3135    | 55                          | 63.3    |
| <i>Aspergillus fumigatus</i> ATCC 13073  | 55                          | 62.5    |
| <i>Aspergillus terreus</i><br>9A-1       | 49                          | 57.5    |
| <i>Aspergillus terreus</i><br>cbs116.46  | 45                          | 58.5    |
| <i>Emmericella nidulans</i>              | 45                          | 55.7    |
| <i>Myceliophthora thermophila</i>        | 55                          | -       |
| <i>Talaromyces thermophilus</i>          | 45                          | -       |
| Consensus phytase-10-thermo[5]-Q50T-K91A | -                           | 90.4    |
| Consensus-phytase-10-thermo[3]-Q50T-K91A | 82                          | 89.3    |

55

|                                                                                         |    |      |
|-----------------------------------------------------------------------------------------|----|------|
| Consensus-phytase-10-thermo [3] -Q50T                                                   | 82 | 88.6 |
| Consensus-phytase-10                                                                    | 80 | 85.4 |
| Consensus phytase-1-thermo [11] -Q50T-K91A                                              | -  | 88.0 |
| Consensus phytase-1-thermo [11] -Q50T                                                   | -  | 88.5 |
| Consensus-phytase-1-thermo [8] -Q50T-K91A                                               | -  | 85.7 |
| Consensus-phytase-1-thermo [8] -Q50T                                                    | 78 | 84.7 |
| Consensus-phytase-1-thermo [8]                                                          | 81 | -    |
| Consensus-phytase-1-thermo [3]                                                          | 75 | -    |
| Consensus-phytase-1-Q50T                                                                | -  | 78.9 |
| Consensus-phytase-1                                                                     | 71 | 78.1 |
| Aspergillus fumigatus $\alpha$ -mutant Q51T                                             | 60 | 67.0 |
| Aspergillus fumigatus $\alpha$ -mutant, plus mutations E59A, S154N, R329H, S364T, G404A | 63 | -    |
| Aspergillus fumigatus "optimized" alpha-mutant, plus mutation K92A                      | 63 | -    |

DQCFD "S333450

Example 10Determination of the melting temperature by differential scanning calorimetry (DSC)

5 In order to determine the unfolding temperature of the phytases, differential scanning calorimetry was applied as described by Brugger et al., 1997 [Brugger, R., Mascarello, F., Augem, S., van Loon, A. P. G. M. & Wyss, M. (1997). Thermal denaturation of phytases and pH 2.5 acid phosphatase studied by  
 10 differential scanning calorimetry. In The Biochemistry of phytate and phytase (eds. Rasmussen, S.K.; Raboy, V.; Dalbøge, H. and Loewus, F.; Kluwer Academic Publishers, Dordrecht, the Netherlands]. Solutions of 50-60 mg/ml of homogeneous phytase were used for the tests. A constant heating rate of 10°C/min was ap-  
 15 plied up to 90-95°C.

The determined melting points confirm the results obtained for the temperature optima (Table 7). The most stable consensus phytase designed so far is consensus phytase-10-thermo[3]-Q50T-K91A showing a melting temperature under the chosen conditions  
 20 of 89.3°C. This is 26.0 to 33.6°C higher than the melting temperature of the wild-type phytases used.

Example 11Transfer of basidiomycete phytase active site into  
 25 consensus phytase-10-thermo[3]-Q50T-K91A

As described previously (Example 5), mutations derived from the basidiomycete phytase active sites were introduced into consensus phytase-10. The following five constructs a) to e) were prepared:

30 a) The construct called consensus phytase-12, and it comprises a selected number of active site residues of the

basidio consensus sequence. Its amino acid sequence is shown in Fig. 21 (the first 26 amino acids form the signal peptide; positions differing from consensus phytase-10-thermo[3]-Q50T-K91A are underlined);

5        b) a cluster of mutations (Cluster II) was transferred to the consensus phytase-1 and -10 sequences, viz.: S80Q, Y86F, S90G, K91A, S92A, K93T, A94R, Y95I;

      c) in a similar way, another cluster of mutations (Cluster III) was transferred, viz.: T129V, E133A, Q134N, M136S, V137S,  
10 N138Q, S139A;

      d) in a similar way, a further cluster of mutations (Cluster IV) was transferred, viz.: A168D, E171T, K172N, F173W;

      e) and finally, a further cluster of mutations (Cluster V) was transferred, viz.: Q297G, S298D, G300D, Y305T.

15        These constructs were expressed as described in Examples 6 to 8.

#### Example 12

##### Phytase alignment using GAP

20        The phytases described herein - i.e. the amino acid sequences as well as the corresponding DNA sequences - were aligned against each other. Also some other phytases were correspondingly aligned, viz. the following:

- the consensus phytase-1 described in EP 897985;
- 25 - the phytase derived from *Aspergillus niger* (ficuum) NRRL 3135 (*A. niger* NRRL3135) described in EP 420358;
- the phytases derived from *Aspergillus fumigatus* ATCC 13073 (*A. fumigatus* 13073); *Aspergillus fumigatus* ATCC 32239 (*A. fumigatus* 32239); *Aspergillus terreus* cbs116.46 (*A. terreus* cbs);
- 30 *Emericella nidulans* (*E. nidulans*); and *Talaromyces thermophilus* (*T. thermophilus*) - all described in EP 897010;

030210 5333450

- the phytases derived from *Myceliophthora thermophila* (*M. thermophila*); and *Aspergillus terreus* 9-A1 (*A. terreus* 9-A1) - both described in EP 684313;
- the phytase derived from *Thermomyces lanuginosus* (*T. lanuginosus*) described in WO 9735017 (PCT/US97/04559);
- the phytases derived from *Agrocybe pediades* (*A. pediades*), *Paxillus involutus* 1 and 2 (*P. involutus* phyA1 and phyA2); and *Trametes pubescens* (*T. pubescens*) - all described in WO 98/28409; and
- 10 - the phytase derived from *Peniophora lycii* (*P. lycii*) described in WO 98/28408.

For the alignments, the program GAP was used with the settings as described above.

For polypeptide comparisons, the signal peptides were  
15 included with the exception of comparisons with consensus phytase-11.

The results of the amino acid sequence comparisons are shown in Table 8 below. The first number in each cell is the amino acid similarity, the second number is the amino acid  
20 identity.

For DNA sequence comparisons, the signal sequence was always included. The results are shown in Table 9 below.

This invention comprises e.g. the following embodiments (A) to (J) that are described below.

25 In these embodiments, when determining % identity or % similarity at the amino acid level for another phytase, its amino acid sequence is aligned with the reference sequence (e.g. in embodiment (A) the consensus phytase-10 amino acid sequence), using an alignment program such as GAP referred to above.  
30 Percentage identity, as well as percentage similarity, is

calculated by the program. The amino acid sequence of the other phytase may or may not include the signal peptide.

When determining % identity on the DNA level for another phytase-encoding DNA, this DNA sequence is aligned with the  
5 reference sequence [e.g. in embodiment (A) nucleotides 12-1412 of SEQ ID NO: 25 (the DNA sequence of consensus phytase-10 (Fcp10) as shown in Fig. 5], using an alignment program such as GAP referred to above. Percentage identity is calculated by the program. The DNA sequence encoding the other phytase can be a  
10 genomic DNA sequence including introns, or it can be a cDNA sequence. It may or may not include the signal peptide-encoding part.

When determining hybridization, the probe to be used is the specified DNA sequence [e.g. in embodiment (A) nucleotides  
15 12-1412 of SEQ ID NO: 25 (the DNA sequence of consensus phytase-10 (Fcp10) as shown in Fig. 5)]. The DNA sequence encoding the other phytase can be a genomic DNA sample which contains a phytase-encoding DNA-sequence; a purified genomic DNA sequence (purified with respect to the phytase-encoding DNA sequence); or  
20 it can be a phytase-encoding cDNA sequence, preferably purified or amplified, e.g. PCR-amplified. The phytase-encoding DNA, whatever type, may or may not include the signal peptide-encoding part. Suitable hybridization conditions are referred to above.

25 The term "DNA sequence" includes such fragments or parts of the herein exemplified DNA sequences, as long as they are capable of encoding an active enzyme (e.g. phytase).

The term "amino acid sequence" includes such fragments or parts of the herein exemplified amino acid sequences, as long  
30 as they are enzymatically active (e.g. displaying phytase activity).

(A) Phytases and corresponding DNA sequences related to consensus phytase-10 (CP10, Fcp 10)

A phytase that comprises an amino acid sequence which is at least 93.80%; or at least 94, 94.5, 95, 95.5, 96, 96.5, 97, 5 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-10 (Fcp10) as shown in Fig. 5.

A phytase that comprises an amino acid sequence which is at least 95.09%; or at least 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% similar to the sequence of amino acids 1-467 of 10 consensus phytase-10.

A phytase which is encoded by a DNA sequence which is at least 95.88%; or at least 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to nucleotides 12-1412 of the DNA sequence of consensus phytase-10 (Fcp10) as shown in Fig. 5.

15 A DNA sequence which encodes a phytase and which (i) is at least 95.88%; or at least 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical; or (ii) hybridizes under low, or medium, medium/high, high, or very high stringency conditions to nucleotides 12-1412 of the DNA sequence of consensus phytase-10 20 (Fcp10) as shown in Fig. 5. A suitable negative control is DNA encoding consensus phytase-1. A suitable positive control is DNA encoding any of CP10, CP10-thermo[3]-Q50T, K91A, CP1-thermo[8], CP1-thermo[8]Q50T,K91A.

A DNA sequence which encodes a phytase comprising an amino 25 acid sequence which is at least 93.80%; or at least 94, 94.5, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-10 (Fcp10) as shown in Fig. 5.

(B) Phytases and corresponding DNA sequences related to  
consensus phytase-10-thermo[3]-Q50T-K91A

A phytase which comprises an amino acid sequence which is at least 93.37%; or at least 93.5, 94, 94.5, 95, 95.5, 96, 96.5, 5 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-10-thermo[3]-Q50T-K91A as shown in Fig. 8.

A phytase which comprises an amino acid sequence which is at least 94.66%; or at least 95.0, 95.5, 96, 96.5, 97, 97.5, 98, 10 98.5, 99, 99.5% similar to the sequence of amino acids 1-467 of consensus phytase-10-thermo[3]-Q50T-K91A as shown in Fig. 8.

A phytase which is encoded by a DNA sequence which is at least 95.88%; or at least 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to nucleotides 12-1412 of the DNA sequence of 15 consensus phytase-10-thermo[3]-Q50T-K91A as shown in Fig. 8.

A DNA sequence which encodes a phytase and which (i) is at least 95.88%; or at least 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical; or (ii) hybridizes under low, or medium, medium/high, high, or very high stringency conditions to 20 nucleotides 12-1412 of the DNA sequence of consensus phytase-10-thermo[3]-Q50T-K91A as shown in Fig. 8. A suitable negative control is DNA encoding consensus phytase-1. A suitable positive control is DNA encoding any of CP10, CP10-thermo[3]-Q50T-K91A, CP1-thermo[8], or CP1-thermo[8]-Q50T-K91A.

25 A DNA sequence which encodes a phytase comprising an amino acid sequence which is at least 93.37%; or at least 93.5, 94, 94.5, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-10-thermo[3]-Q50T-K91A as shown in Fig. 8.

(C) Phytases and corresponding DNA sequences related to consensus phytase-1-thermo[8]

A phytase which comprises an amino acid sequence which is at least 98.30%; or at least 98.5, 99, 99.5% identical to the  
5 sequence of amino acids 1-467 of consensus phytase-1-thermo[8] (as shown in Fig. 7; backmutations T50Q and A91K to be added).

A phytase which comprises an amino acid sequence which is at least 98.51%; or at least 99, 99.5% similar to the sequence of amino acids 1-467 of consensus phytase-1-thermo[8] (as shown  
10 in Fig. 7; backmutations T50Q and A91K to be added).

A phytase which is encoded by a DNA sequence which is at least 98.73%; or at least 99, 99.5% identical to nucleotides 1-1407 of the DNA sequence of consensus phytase-1-thermo[8] (as shown in Fig. 7; backmutations T50Q and A91K to be added).

15 A DNA sequence which encodes a phytase and which (i) is at least 98.73%; or at least 99, 99.5% identical; or (ii) hybridizes under low, or medium, medium/high, high, or very high stringency conditions to nucleotides 1-1407 of the DNA sequence of consensus phytase-1-thermo[8] (as shown in Fig. 7;  
20 backmutations T50Q and A91K to be added). A suitable negative control is DNA encoding consensus phytase-1. A suitable positive control is DNA encoding any of CP1-thermo[8], CP1-thermo[8]-Q50T-K91A.

A DNA sequence which encodes a phytase comprising an amino  
25 acid sequence which is at least 98.30%; or at least 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-1-thermo[8] (as shown in Fig. 7; backmutations T50Q and A91K to be added).

(D) Phytases and corresponding DNA sequences related to  
consensus phytase-1-thermo[8]-Q50T-K91A

A phytase which comprises an amino acid sequence which is at least 97.87%; or at least 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-1-thermo[8]-Q50T-K91A as shown in Fig. 7.

A phytase which comprises an amino acid sequence which is at least 98.08%; or at least 98.5, 99, 99.5% similar to the sequence of amino acids 1-467 of consensus phytase-1-thermo[8]-Q50T-K91A as shown in Fig. 7.

A phytase which is encoded by a DNA sequence which is at least 98.37%; or at least 98.5, 99, 99.5% identical to nucleotides 1-1407 of the DNA sequence of consensus phytase-1-thermo[8]-Q50T-K91A as shown in Fig. 7.

A DNA sequence which encodes a phytase and which (i) is at least 98.37%; or at least 98.5, 99, 99.5% identical; or (ii) hybridizes under low, or medium, medium/high, high, or very high stringency conditions to nucleotides 1-1407 of the DNA sequence of consensus phytase-1-thermo[8]-Q50T-K91A as shown in Fig. 7. A suitable negative control is DNA encoding consensus phytase-1. A suitable positive control is DNA encoding any of CP1-thermo[8], CP1-thermo[8]-Q50T-K91A.

A DNA sequence which encodes a phytase comprising an amino acid sequence which is at least 97.87%; or at least 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase 1-thermo[8]-Q50T-K91A as shown in Fig. 7.

(E) Phytases and corresponding DNA sequences related to  
consensus phytase-11

A phytase that comprises an amino acid sequence which is at least 90.71%; or at least 91, 91.5, 92, 92.5, 93, 93.5, 94,

94.5, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-482 of consensus phytase-11 as shown in Fig. 6.

A phytase that comprises an amino acid sequence which is at least 92.07%; or at least 92.5, 93, 93.5, 94, 94.5, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% similar to the sequence of amino acids 1-482 of consensus phytase-11 as shown in Fig. 6.

A DNA sequence that encodes a phytase comprising an amino acid sequence which is at least 90.71%; or at least 91, 91.5, 92, 92.5, 93, 93.5, 94, 94.5, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-482 of consensus phytase-11 as shown in Fig. 6.

(F) Phytases and corresponding DNA sequences related to A. fumigatus alpha-mutant

A phytase that comprises an amino acid sequence which is at least 97.17%; or at least 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of A. fumigatus alpha-mutant (phytase) as shown in Fig. 9.

A phytase that comprises an amino acid sequence that is at least 97.82%; or at least 98, 98.5, 99, 99.5% similar to the sequence of amino acids 1-467 of A. fumigatus alpha-mutant (phytase) as shown in Fig. 9.

A phytase which is encoded by a DNA sequence which is at least 96.13%; or at least 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to nucleotides 1-1401 of the DNA sequence of A. fumigatus ATCC 13073 alpha-mutant (phytase) as shown in Fig. 9.

A DNA sequence which encodes a phytase comprising an amino acid sequence which is at least 97.17%; or at least 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467

of *A. fumigatus* ATCC 13073 alpha-mutant (phytase) as shown in Fig. 9.

A DNA sequence which encodes a phytase and which (i) is at least 96.13%; or 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical; or (ii) hybridizes under low, or medium, medium/high, high, or very high stringency conditions to nucleotides 1-1401 of the DNA sequence of *A. fumigatus* ATCC 13073 alpha-mutant (phytase) shown in Fig. 9. A suitable negative control is DNA encoding *A. fumigatus* 13073 phytase. A suitable positive control is DNA encoding any one of the *A. fumigatus* ATCC 13073 alpha mutant phytase or the optimised alpha-mutant.

(G) Phytases and corresponding DNA sequences related to the optimized *A. fumigatus* alpha-mutant

A phytase that comprises an amino acid sequence that is at least 96.08%; or at least 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of the phytase of the optimized *A. fumigatus* alpha-mutant.

A phytase that comprises an amino acid sequence that is at least 96.74%; or at least 97, 97.5, 98, 98.5, 99, 99.5% similar to the sequence of the phytase of the optimized *A. fumigatus* alpha-mutant.

A phytase which is encoded by a DNA sequence which is at least 95.63%; or at least 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to nucleotides 1-1401 of the DNA sequence encoding the optimized *A. fumigatus* alpha-mutant phytase.

A DNA sequence that encodes a phytase comprising an amino acid sequence that is at least 96.08%; or at least 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the optimized *A. fumigatus* alpha-mutant phytase.

A DNA sequence which encodes a phytase and which (i) is at least 95.63%; or at least 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical; or (ii) hybridizes under low, or medium, medium/high, high, very high stringency conditions to 5 nucleotides 1-1401 of the DNA sequence encoding the optimized *A. fumigatus* alpha-mutant phytase.

A suitable negative control is DNA encoding *A. fumigatus* ATCC 13073 phytase. A suitable positive control is DNA encoding any one of the *A. fumigatus* ATCC 13073 alpha mutant phytase of 10 the optimised alpha-mutant.

(H) Phytases and corresponding DNA sequences related to consensus phytase-7

A phytase that comprises an amino acid sequence which is 15 at least 94.87%; or at least 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-7 as shown in Fig. 10.

A phytase that comprises an amino acid sequence which is at least 95.30%; or at least 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 20 99, 99.5% similar to the sequence of amino acids 1-467 of consensus phytase-7 as shown in Fig. 10.

A phytase which is encoded by a DNA sequence which is at least 96.38%; or 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to nucleotides 12-1412 of the DNA sequence of consensus phytase- 25 7 shown in Fig. 10.

A DNA sequence which encodes a phytase and which (i) is at least 96.38%; or at least 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical; or (ii) hybridizes under low, or medium, medium/high, high, or very high stringency conditions to nucleotides 12-1412 30 of the DNA sequence of consensus phytase-7 as shown in Fig. 10.

000210" 59233460

A DNA sequence which encodes a phytase comprising an amino acid sequence which is at least 94.87%; or at least 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-7 as shown in 5 Fig. 10.

(I) Phytases related to basidio consensus phytase

A phytase which comprises an amino acid sequence which is at least 76.23%; or at least 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 94.5, 95, 95.5, 96, 96.5, 10 97, 97.5, 98, 98.5, 99, 99.5% identical to the combined sequence of (i) amino acids 1-441 of basidio consensus phytase shown in Fig. 3, and (ii) amino acids 1-26 shown in Fig. 5 (the sequence of (ii) to be added at the N-terminus of the sequence of (i)).

A phytase that comprises an amino acid sequence which is 15 at least 79.50%; or at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% similar to the sequence of amino acids 1-441 of basidio consensus phytase as shown in Fig. 3.

(J) Phytases related to consensus phytase-12

20 A phytase which comprises an amino acid sequence which is at least 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% identical to the sequence of amino acids 1-467 of consensus phytase-12 as shown in Fig. 21.

25 A phytase which comprises an amino acid sequence which is at least 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95.5, 96, 96.5, 97, 97.5, 98, 98.5, 99, 99.5% similar to the sequence of amino acids 1-467 of consensus phytase-12 as shown in Fig. 21.

DocId: 59232450

Table 8

## Comparison of phytase amino acid sequences

| Phytase             | CP10        | CP10-thermo [3] - Q50T-K91A | CP1-thermo [8] | CP1-thermo [8] - Q50T-K91A | CP11        | CP7         | Basidio     | A. fumigatus alpha-mutant | A. fumigatus alpha-mutant (opt.) |
|---------------------|-------------|-----------------------------|----------------|----------------------------|-------------|-------------|-------------|---------------------------|----------------------------------|
| Consensus phytase-1 | 95.08/93.79 | 94.65/93.36                 | 98.50/98.29    | 98.07/97.86                | 92.06/90.70 | 95.29/94.86 | 69.42/62.16 | 85.59/82.58               | 84.73/81.72                      |
| A. niger NRRL3135   | 79.48/76.46 | 79.05/76.03                 | 80.35/77.75    | 79.91/77.32                | 79.27/76.31 | 84.02/81.64 | 67.19/59.32 | 74.07/70.11               | 74.95/70.99                      |
| A. terreus 9-A1     | 76.04/72.11 | 75.82/71.90                 | 76.47/72.33    | 76.25/72.11                | 76.51/73.02 | 75.76/71.18 | 65.39/58.02 | 69.67/64.84               | 69.45/64.84                      |
| A. terreus cbs      | 79.04/75.11 | 78.82/74.89                 | 79.48/75.76    | 79.26/75.55                | 77.19/73.27 | 79.17/75.00 | 66.92/59.65 | 72.59/67.76               | 72.37/67.76                      |
| E. nidulans         | 78.70/74.35 | 78.26/73.91                 | 79.78/75.87    | 79.35/75.44                | 80.56/76.62 | 76.96/73.04 | 67.20/58.13 | 72.39/67.83               | 72.11/67.54                      |
| A. fumigatus 13073  | 82.93/80.31 | 82.50/79.87                 | 82.31/79.04    | 81.88/78.60                | 81.36/78.64 | 80.13/76.20 | 63.54/57.91 | 97.82/97.16               | 96.73/96.07                      |
| A. fumigatus 32239  | 81.30/77.39 | 80.87/76.96                 | 81.09/77.61    | 80.65/77.17                | 79.95/76.08 | 79.13/75.22 | 63.61/54.97 | 90.22/86.52               | 89.57/85.87                      |
| T. thermophilus     | 77.83/73.84 | 77.38/73.39                 | 78.67/74.89    | 78.22/74.44                | 78.47/74.76 | 76.51/73.15 | 61.54/54.36 | 72.01/66.82               | 72.69/67.49                      |
| M. thermophila      | 69.16/62.81 | 69.48/63.33                 | 69.27/62.84    | 69.59/63.36                | 69.65/63.06 | 68.82/62.13 | 65.56/57.91 | 66.21/58.45               | 66.44/58.68                      |
| T. lanuginosus      | 73.52/66.70 | 73.06/66.44                 | 71.92/64.61    | 71.46/64.16                | 74.21/68.86 | 69.50/62.62 | 67.20/57.41 | 68.91/61.02               | 69.61/61.72                      |
| P. lycii            | 64.92/59.10 | 64.91/59.37                 | 64.46/58.09    | 64.46/58.36                | 65.03/59.84 | 63.13/56.50 | 77.75/73.07 | 64.08/57.11               | 62.47/55.91                      |
| A. pediades         | 64.51/51.81 | 64.86/51.94                 | 62.98/51.41    | 63.33/51.54                | 64.50/52.30 | 63.05/51.15 | 78.92/74.71 | 61.64/52.38               | 62.13/53.07                      |
| P. involutus 1      | 66.67/58.07 | 66.67/58.33                 | 64.84/56.51    | 64.84/56.77                | 63.30/54.52 | 65.33/56.53 | 79.49/76.22 | 59.59/51.81               | 59.95/52.20                      |
| P. involutus 2      | 65.54/55.70 | 65.30/55.53                 | 66.85/56.87    | 66.58/56.68                | 66.30/56.35 | 64.27/54.13 | 78.09/74.59 | 61.26/52.62               | 61.04/52.47                      |

|                         |             |             |             |             |             |             |             |             |             |
|-------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| T. pubescens            | 65.46/57.22 | 65.72/57.47 | 62.89/55.67 | 63.14/55.93 | 65.03/57.65 | 63.28/56.51 | 78.34/75.12 | 64.08/57.11 | 62.30/55.24 |
| CP10                    | -           | 99.57/99.57 | 96.57/95.50 | 96.15/95.08 | 95.02/94.56 | 91.01/89.29 | 70.22/62.28 | 85.13/82.76 | 85.99/83.62 |
| CP10t [3]Q50TK91A       | 99.57/99.57 | -           | 96.15/95.08 | 96.57/95.50 | 94.56/94.10 | 90.58/88.87 | 70.47/62.28 | 85.13/82.76 | 85.99/83.62 |
| CP1Thermo [8]           | 96.57/95.50 | 96.15/95.08 | -           | 99.57/99.57 | 93.42/92.29 | 94.43/93.79 | 68.40/60.74 | 84.52/81.94 | 85.38/82.80 |
| CP1t [8]Q50TK91A        | 96.15/95.08 | 96.57/95.50 | 99.57/99.57 | -           | 92.97/91.84 | 94.00/93.36 | 68.64/60.74 | 84.52/81.94 | 85.38/82.80 |
| CP11                    | 95.02/94.56 | 94.56/94.10 | 93.42/92.29 | 92.97/91.84 | -           | 88.44/86.62 | 68.27/59.73 | 82.23/79.73 | 83.37/80.87 |
| CP7                     | 91.01/89.29 | 90.58/88.87 | 94.43/93.79 | 94.00/93.36 | 88.44/86.62 | -           | 69.80/62.69 | 81.94/78.71 | 81.72/78.50 |
| Basidio                 | 70.22/62.28 | 70.47/62.28 | 68.40/60.74 | 68.64/60.74 | 68.27/59.73 | 69.80/62.69 | -           | 65.97/60.52 | 66.41/60.68 |
| A. fumigatus alpha-mut. | 85.13/82.76 | 85.13/82.76 | 84.52/81.94 | 84.52/81.94 | 82.23/79.73 | 81.94/78.71 | 65.97/60.52 | -           | 98.93/98.93 |
| A. fum alpha-mut - opt. | 85.99/83.62 | 85.99/83.62 | 85.38/82.80 | 85.38/82.80 | 83.37/80.87 | 81.72/78.50 | 66.41/60.68 | 98.93/98.93 | -           |

Table 9

## Comparison of phytase encoding DNA sequences

| Phytase             | CP10  | CP10-thermo[3] - Q50T-K91A | CP1-thermo[8] | CP1-thermo[8] - Q50T-K91A | CP7   | Basidio | A. fumigatus alpha-mutant | A. fumi-gatus alpha-mutant (opt.) |
|---------------------|-------|----------------------------|---------------|---------------------------|-------|---------|---------------------------|-----------------------------------|
| Consensus phytase-1 | 95.87 | 95.87                      | 98.72         | 98.36                     | 96.37 | 65.46   | 66.88                     | 66.88                             |
| A. niger NRRL3135   | 65.10 | 64.82                      | 66.10         | 65.74                     | 67.52 | 50.68   | 65.88                     | 66.17                             |
| A. terreus 9-A1     | 61.74 | 61.53                      | 62.17         | 62.03                     | 60.53 | 49.40   | 66.24                     | 66.31                             |
| A. terreus cbs      | 62.52 | 62.30                      | 63.02         | 62.88                     | 61.45 | 49.74   | 68.17                     | 68.24                             |
| E. nidulans         | 65.08 | 64.94                      | 65.30         | 65.01                     | 64.22 | 49.92   | 64.90                     | 65.44                             |
| A. fumigatus 13073  | 65.66 | 65.38                      | 64.19         | 64.08                     | 63.65 | 48.27   | 96.12                     | 95.62                             |
| T. thermophilus     | 62.52 | 62.50                      | 62.53         | 62.66                     | 62.00 | 52.19   | 61.77                     | 61.92                             |
| M. thermophila      | 55.51 | 55.15                      | 55.36         | 55.22                     | 53.91 | 48.44   | 58.17                     | 58.24                             |
| T. lanuginosus      | 57.56 | 57.20                      | 56.76         | 56.47                     | 62.00 | 44.66   | 59.71                     | 60.07                             |
| P. lycii            | 45.76 | 46.51                      | 45.14         | 55.21                     | 55.46 | 58.50   | 48.91                     | 49.44                             |
| A. pediades         | 49.89 | 49.89                      | 49.89         | 50.11                     | 45.54 | 61.66   | 47.49                     | 47.56                             |
| P. involutus 1      | 48.32 | 49.03                      | 47.81         | 47.96                     | 49.59 | 59.80   | 49.96                     | 50.19                             |
| P. involutus 2      | 48.24 | 49.00                      | 48.08         | 48.63                     | 47.94 | 60.16   | 47.56                     | 47.63                             |
| T. pubescens        | 47.00 | 47.17                      | 46.46         | 47.62                     | 46.83 | 60.37   | 49.89                     | 49.96                             |
| CP10                | -     | 99.43                      | 96.40         | 96.05                     | 93.73 | 66.40   | 67.81                     | 68.24                             |

|                         |       |       |       |       |       |       |       |       |
|-------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| CP10T[3] Q50TK91A       | 99.43 | -     | 96.37 | 96.58 | 93.45 | 66.29 | 67.81 | 68.24 |
| Cp1thermo[8]            | 96.40 | 96.37 | -     | 99.65 | 95.30 | 65.40 | 66.74 | 67.17 |
| CP1t[8] Q50TK91A        | 96.05 | 96.58 | 99.65 | -     | 94.94 | 65.47 | 66.74 | 67.17 |
| CP7                     | 93.73 | 93.45 | 95.30 | 94.94 | -     | 64.56 | 65.88 | 65.88 |
| Basidio                 | 66.40 | 66.29 | 65.40 | 65.47 | 64.56 | -     | 50.41 | 50.49 |
| A. fumigatus alpha-mut. | 67.81 | 67.81 | 66.74 | 66.74 | 65.88 | 50.41 | -     | 99.50 |
| A. fum alpha-mut - opt. | 68.24 | 68.24 | 67.17 | 67.17 | 65.88 | 50.49 | 99.50 | -     |



- (ii) variants of (i), further including the mutations Q50T, K91A, or (Q50T+K91A), the latter variant being shown in Fig. 7,
- (iii) amino acids 27-467 of any of the sequences of (i) and (ii), or
- (iv) SEQ ID NO: 29, or amino acids 1-441 thereof; or an amino acid sequence encoded by
- (v) nucleotides 1-1407, or 79-1407 of SEQ ID NO: 28.
6. A phytase that comprises the amino acid sequence of consensus phytase-11 (SEQ ID NO: 27).
7. A DNA sequence that comprises a DNA-sequence encoding a phytase of claim 1.
8. A DNA sequence that comprises a DNA-sequence encoding a phytase, wherein the phytase-encoding DNA-sequence is (i) at least 95.88% identical, or (ii) hybridizes under high stringency conditions, to nucleotides 12-1412 of the DNA sequence of consensus phytase-10 (SEQ ID NO: 25).
9. A DNA sequence that comprises a DNA-sequence that encodes a phytase, wherein the phytase comprises an amino acid sequence that is at least 93.80% identical to the sequence of amino acids 1-467 of consensus phytase 10 (SEQ ID NO: 26).
10. A DNA sequence that comprises a DNA-sequence that encodes a phytase, and wherein the phytase-encoding DNA-sequence comprises
- (i) nucleotides 12-1412, or 90-1412 of the DNA sequence of consensus phytase-10 (SEQ ID NO: 25);

- (ii) nucleotides 1-1401, or 79-1401 of the DNA sequence of consensus phytase-10-thermo[3]-Q50T-K91A (SEQ ID NO: 30); or
- (iii) nucleotides 1-1401, or 79-1401 of the DNA sequence of consensus phytase-1-thermo[8]-Q50T-K91A (SEQ ID NO: 28).

11. A vector comprising a DNA sequence according to claim 7.
12. A microbial host cell comprising a DNA sequence according to claim 7, or the vector according to claim 11.
13. A process for producing a phytase, the process comprising culturing the host cell according to claim 12 under conditions permitting the production of the phytase, and recovering the phytase from the culture broth.
14. A food, feed or pharmaceutical composition comprising a phytase of claim 1.

**ABSTRACT**

This invention relates to improved phytases, preferably phytases of an increased thermostability, and a process of producing them. In particular, stabilizing amino acid mutations  
5 are introduced into a homologous protein, or the active site of a phytase is replaced in part or in toto. The corresponding DNA sequences and methods of preparing them are also disclosed, as are methods of producing the improved phytases, and the use thereof. Specific variants of *Aspergillus fumigatus* phytase and  
10 of consensus phytases are disclosed.

030310 "030310"

|                                     | 1          |            |            |            | 50         |
|-------------------------------------|------------|------------|------------|------------|------------|
| <i>A. terreus</i> 9A-1              | KhsDCNSVDh | GYQCFPELSH | kwGLYAPYFS | LQDESPFPID | VPEDChITFV |
| <i>A. terreus</i> cbs               | NhsDCTSVDr | GYQCFPELSH | kwGLYAPYFS | LQDESPFPID | VPDDChITFV |
| <i>A. niger</i> var. <i>awamori</i> | NqsTCDTVDQ | GYQCFSETSH | LWGQYAPFFS | LANESAISPD | VPAGCrVTFA |
| <i>A. niger</i> T213                | NqsSCDTVDQ | GYQCFSETSH | LWGQYAPFFS | LANESVISPD | VPAGCrVTFA |
| <i>A. niger</i> NRRL3135            | NqsSCDTVDQ | GYQCFSETSH | LWGQYAPFFS | LANESVISPE | VPAGCrVTFA |
| <i>A. fumigatus</i> 13073           | GSkSCDTVDI | GYQCSPATSH | LWGQYSPFFS | LEDELSVSSK | LPKDCrITLV |
| <i>A. fumigatus</i> 32722           | GSkSCDTVDI | GYQCSPATSH | LWGQYSPFFS | LEDELSVSSK | LPKDCrITLV |
| <i>A. fumigatus</i> 58128           | GSkSCDTVDI | GYQCSPATSH | LWGQYSPFFS | LEDELSVSSK | LPKDCrITLV |
| <i>A. fumigatus</i> 26906           | GSkSCDTVDI | GYQCSPATSH | LWGQYSPFFS | LEDELSVSSK | LPKDCrITLV |
| <i>A. fumigatus</i> 32239           | GSkACDTVEI | GYQCSPGTSH | LWGQYSPFFS | LEDELSVSSD | LPKDCrVTFV |
| <i>E. nidulans</i>                  | QNHSCNTADG | GYQCFPNVSH | VWGQYSPYFS | IEQESAISeD | VPHGCEvTFV |
| <i>T. thermophilus</i>              | DSHSCNTVEG | GYQCrPEISH | swGQYSPFFS | LADQSEISPD | VPQNCKITFV |
| <i>M. thermophila</i>               | ESRPCDTpDI | GFQCGTAISH | FWGQYSPYFS | VpSElDaS.. | IPDDCEvTFA |
| Consensus                           | NSHSCDTVDG | GYQCFPEISH | LWGQYSPYFS | LEDESAISPD | VPDDC-VTFV |
| Consensus phytase                   | NSHSCDTVDG | GYQCFPEISH | LWGQYSPYFS | LEDESAISPD | VPDDCRVTFV |

|                                     | 51         |            |            |            | 100         |
|-------------------------------------|------------|------------|------------|------------|-------------|
| <i>A. terreus</i> 9A-1              | QVLARHGARS | PTHSKtKAYA | AtIAAIQKSA | TaFpGKYAFL | QSYNYSLDSE  |
| <i>A. terreus</i> cbs               | QVLARHGARS | PTDSKtKAYA | AtIAAIQKNA | TaLpGKYAFL | KSYNYSMGSE  |
| <i>A. niger</i> var. <i>awamori</i> | QVLSRHGARY | PTESKsKkYS | ALIEEIQQNV | TtFDGKYAFL | KTYNYSLGAD  |
| <i>A. niger</i> T213                | QVLSRHGARY | PTESKsKkYS | ALIEEIQQNV | TtFDGKYAFL | KTYNYSLGAD  |
| <i>A. niger</i> NRRL3135            | QVLSRHGARY | PTDSKsKkYS | ALIEEIQQNA | TtFDGKYAFL | KTYNYSLGAD  |
| <i>A. fumigatus</i> 13073           | QVLSRHGARY | PTSSKsKkYK | kLVTAIQaNA | TdFKGKFAFL | KTYNYTLGAD  |
| <i>A. fumigatus</i> 32722           | QVLSRHGARY | PTSSKsKkYK | kLVTAIQaNA | TdFKGKFAFL | KTYNYTLGAD  |
| <i>A. fumigatus</i> 58128           | QVLSRHGARY | PTSSKsKkYK | kLVTAIQaNA | TdFKGKFAFL | KTYNYTLGAD  |
| <i>A. fumigatus</i> 26906           | QVLSRHGARY | PTSSKsKkYK | kLVTAIQaNA | TdFKGKFAFL | KTYNYTLGAD  |
| <i>A. fumigatus</i> 32239           | QVLSRHGARY | PTASKsKkYK | kLVTAIQKNA | TeFKGKFAFL | ETNYNYTLGAD |
| <i>E. nidulans</i>                  | QVLSRHGARY | PTESKsKAYS | GLIEAIQKNA | TsFwGQYAFL | ESYNYTLGAD  |
| <i>T. thermophilus</i>              | QLLSRHGARY | PTSSKtELYS | QLISrIQKTA | TaYKGyYAFL | KDYrYqLGAN  |
| <i>M. thermophila</i>               | QVLSRHGARA | PTlKRaasYv | DLIDrIHhGA | IsYgPgYEFL | RTYDYTLGAD  |
| Consensus                           | QVLSRHGARY | PTSSK-KAYS | ALIEAIQKNA | T-FKGKYAFL | KTYNYTLGAD  |
| Consensus phytase                   | QVLSRHGARY | PTSSKSKAYS | ALIEAIQKNA | TAFKGKYAFL | KTYNYTLGAD  |

|                                     | 101        |            |            |            | 150        |
|-------------------------------------|------------|------------|------------|------------|------------|
| <i>A. terreus</i> 9A-1              | ELTPFGrNQL | rDlGaQFYeR | YNALTRhInP | FVRATDASRV | hESAeKFVEG |
| <i>A. terreus</i> cbs               | NLTPFGrNQL | qDlGaQFYRR | YDTLTRhInP | FVRAADSSRV | hESAeKFVEG |
| <i>A. niger</i> var. <i>awamori</i> | DLTPFGEQEL | VNSGIKFYQR | YESLTRNIIP | FIRSSGSSRV | IASGEKFIEG |
| <i>A. niger</i> T213                | DLTPFGEQEL | VNSGIKFYQR | YESLTRNIIP | FIRSSGSSRV | IASGEKFIEG |
| <i>A. niger</i> NRRL3135            | DLTPFGEQEL | VNSGIKFYQR | YESLTRNIVP | FIRSSGSSRV | IASGKKFIEG |
| <i>A. fumigatus</i> 13073           | DLTPFGEQQL | VNSGIKFYQR | YKALARSVVP | FIRASGSDRV | IASGEKFIEG |
| <i>A. fumigatus</i> 32722           | DLTPFGEQQL | VNSGIKFYQR | YKALARSVVP | FIRASGSDRV | IASGEKFIEG |
| <i>A. fumigatus</i> 58128           | DLTPFGEQQL | VNSGIKFYQR | YKALARSVVP | FIRASGSDRV | IASGEKFIEG |
| <i>A. fumigatus</i> 26906           | DLTAFGEQQL | VNSGIKFYQR | YKALARSVVP | FIRASGSDRV | IASGEKFIEG |
| <i>A. fumigatus</i> 32239           | DLTPFGEQQM | VNSGIKFYQK | YKALAgSVVP | FIRSSGSDRV | IASGEKFIEG |
| <i>E. nidulans</i>                  | DLTiFGENQM | VDSGaKFYRR | YKNLARKnTP | FIRASGSDRV | VASAeKFING |
| <i>T. thermophilus</i>              | DLTPFGENQM | IQlGIKFYnH | YKSLARNAvP | FVRCSGSDRV | IASGrLFIEG |
| <i>M. thermophila</i>               | ELTRtGQQQM | VNSGIKFYRR | YRALARKsIP | FVRTAGqDRV | VhSAENFTQG |
| Consensus                           | DLTPFGENQM | VNSGIKFYRR | YKALARK-VP | FVRASGSDRV | IASAEKFIEG |
| Consensus phytase                   | DLTPFGENQM | VNSGIKFYRR | YKALARKIVP | FIRASGSDRV | IASAEKFIEG |

Fig. 1a

|                                     |            |            |                                  |
|-------------------------------------|------------|------------|----------------------------------|
|                                     | 151        |            | 200                              |
| <i>A. terreus</i> 9A-1              | FQTARqDDHh | ANpHQSPPrV | DVaIPEGSAY NNTLEHSIcT AFES...STV |
| <i>A. terreus</i> cbs               | FQNARqGDPh | ANpHQSPPrV | DVVIPEGTAY NNTLEHSIcT AFEA...STV |
| <i>A. niger</i> var. <i>awamori</i> | FQSTKLkDPr | AqpgQSSPkI | DVVISeASSs NNTLDPGTCT VFED...SEL |
| <i>A. niger</i> T213                | FQSTKLkDPr | AqpgQSSPkI | DVVISeASSs NNTLDPGTCT VFED...SEL |
| <i>A. niger</i> NRRL3135            | FQSTKLkDPr | AqpgQSSPkI | DVVISeASSs NNTLDPGTCT VFED...SEL |
| <i>A. fumigatus</i> 13073           | FQqAKLADPG | A.TNRAAPAI | SVIIPeSETF NNTLDHGvCT kFEA...SQL |
| <i>A. fumigatus</i> 32722           | FQqAKLADPG | A.TNRAAPAI | SVIIPeSETF NNTLDHGvCT kFEA...SQL |
| <i>A. fumigatus</i> 58128           | FQqAKLADPG | A.TNRAAPAI | SVIIPeSETF NNTLDHGvCT kFEA...SQL |
| <i>A. fumigatus</i> 26906           | FQqAKLADPG | A.TNRAAPAI | SVIIPeSETF NNTLDHGvCT kFEA...SQL |
| <i>A. fumigatus</i> 32239           | FQqANVADPG | A.TNRAAPVI | SVIIPeSETY NNTLDHsvCT NFEA...SEL |
| <i>E. nidulans</i>                  | FRKAQLhDHG | S..gQATPVV | NVIIPeIDGF NNTLDHSTCV SFEN...DEr |
| <i>T. thermophilus</i>              | FQSAKVlDPh | SDkHDAPPTI | NVIIeEGPSY NNTLDtGSCP VFED...SSg |
| <i>M. thermophila</i>               | FHSALLADRG | STvRPTlPyd | mVVIPEtAGa NNTLHNDlCT AFEEgpySTI |
| Consensus                           | FQSAKLADPG | S-PHQASPVI | NVIIPeGSY NNTLDHGtCT AFED---SEL  |
| Consensus phytase                   | FQSAKLADPG | SQPHQASPVI | DVIIPeGSY NNTLDHGtCT AFED...SEL  |

|                                     |            |            |                                  |
|-------------------------------------|------------|------------|----------------------------------|
|                                     | 201        |            | 250                              |
| <i>A. terreus</i> 9A-1              | GDDAVANFTA | VFAPAiaQRL | EADLPgVqLS TDDVVnLMAM CPFETVSItd |
| <i>A. terreus</i> cbs               | GDAAADNFTA | VFAPAiaKRL | EADLPgVqLS ADDVVnLMAM CPFETVSItd |
| <i>A. niger</i> var. <i>awamori</i> | ADTVEANFTA | TFAPSIrQRL | ENDLSGvTLT DTEVTyLMdM CSFDTIstST |
| <i>A. niger</i> T213                | ADTVEANFTA | TFAPSIrQRL | ENDLSGvTLT DTEVTyLMdM CSFDTIstST |
| <i>A. niger</i> NRRL3135            | ADTVEANFTA | TFVPSIrQRL | ENDLSGvTLT DTEVTyLMdM CSFDTIstST |
| <i>A. fumigatus</i> 13073           | GDEVAANFTA | lFAPDIRARa | EkHLPGvTLT DEDVVslMDM CSFDTVARTS |
| <i>A. fumigatus</i> 32722           | GDEVAANFTA | lFAPDIRARa | EkHLPGvTLT DEDVVslMDM CSFDTVARTS |
| <i>A. fumigatus</i> 58128           | GDEVAANFTA | lFAPDIRARa | EkHLPGvTLT DEDVVslMDM CSFDTVARTS |
| <i>A. fumigatus</i> 26906           | GDEVAANFTA | lFAPDIRARa | EkHLPGvTLT DEDVVslMDM CSFDTVARTS |
| <i>A. fumigatus</i> 32239           | GDEVEANFTA | lFAPAIrARI | EkHLPGVqLT DDDVVslMDM CSFDTVARTa |
| <i>E. nidulans</i>                  | ADEiEANFTA | IMGPPIrKRL | ENDLPgIKLT NENViyLMdM CSFDTMARTa |
| <i>T. thermophilus</i>              | GHDAQEKFAk | qFAPAIlEKI | KDHLPGVDLa vSDVpyLMdL CPFETLARNh |
| <i>M. thermophila</i>               | GDDAQDTyLS | TFAGPItARV | NANLPgANLT DADTVaLMdL CPFETVASSS |
| Consensus                           | GDDAEANFTA | TFAPAIrARL | EADLPgVTLT DEDVV-LMDM CPFETVARTS |
| Consensus phytase                   | GDDVEANFTA | LFAPAIrARL | EADLPgVTLT DEDVVyLMdM CPFETVARTS |

|                                     |            |               |                                  |
|-------------------------------------|------------|---------------|----------------------------------|
|                                     | 251        |               | 300                              |
| <i>A. terreus</i> 9A-1              | .....      | ...DAhTLSPFC  | DLFTAtEWtq YNYLlSLDKY YGYGGGNPLG |
| <i>A. terreus</i> cbs               | .....      | ...DAhTLSPFC  | DLFTAaEWtq YNYLlSLDKY YGYGGGNPLG |
| <i>A. niger</i> var. <i>awamori</i> | .....      | ...vDTKLSPFC  | DLFTHdEWih YDYlQSLkKY YGHGAGNPLG |
| <i>A. niger</i> T213                | .....      | ...vDTKLSPFC  | DLFTHdEWih YDYlRSLkKY YGHGAGNPLG |
| <i>A. niger</i> NRRL3135            | .....      | ...vDTKLSPFC  | DLFTHdEWin YDYlQSLkKY YGHGAGNPLG |
| <i>A. fumigatus</i> 13073           | .....      | ...DASQLSPFC  | QLFTHnEWkk YNYlQSLGKY YGYGAGNPLG |
| <i>A. fumigatus</i> 32722           | .....      | ...DASQLSPFC  | QLFTHnEWkk YNYlQSLGKY YGYGAGNPLG |
| <i>A. fumigatus</i> 58128           | .....      | ...DASQLSPFC  | QLFTHnEWkk YNYlQSLGKY YGYGAGNPLG |
| <i>A. fumigatus</i> 26906           | .....      | ...DASQLSPFC  | QLFTHnEWkk YNYlQSLGKY YGYGAGNPLG |
| <i>A. fumigatus</i> 32239           | .....      | ...DASELSPFC  | AlFTHnEWkk YDYlQSLGKY YGYGAGNPLG |
| <i>E. nidulans</i>                  | .....      | ...HGTELSPFC  | AlFTEkEWlq YDYlQSLSKY YGYGAGSPLG |
| <i>T. thermophilus</i>              | .....      | ...TDT..LSPFC | ALSTQeEWqa YDYlQSLGKY YGnGGGNPLG |
| <i>M. thermophila</i>               | sdpatadagg | gNGRpLSPFC    | rLFSEsEWra YDYlQSVGKW YGYGPNPLG  |
| Consensus                           | -----      | -DATELSPFC    | ALFTE-EW-- YDYlQSLGKY YGYGAGNPLG |
| Consensus phytase                   | .....      | ..DATELSPFC   | ALFTHDEWRQ YDYlQSLGKY YGYGAGNPLG |

Fig. 1b

|                                     |            |            |            |            |            |
|-------------------------------------|------------|------------|------------|------------|------------|
|                                     | 301        |            |            |            | 350        |
| <i>A. terreus</i> 9A-1              | PVQGVGWaNE | LMARLTRAPV | HDHTCVNNTL | DASPATFPLN | ATLYADFSHD |
| <i>A. terreus</i> cbs               | PVQGVGWaNE | LIARLTRSPV | HDHTCVNNTL | DANPATFPLN | ATLYADFSHD |
| <i>A. niger</i> var. <i>awamori</i> | PTQGVGYaNE | LIARLTHSPV | HDDTSSNHTL | DSNPATFPLN | STLYADFSHD |
| <i>A. niger</i> T213                | PTQGVGYaNE | LIARLTHSPV | HDDTSSNHTL | DSNPATFPLN | STLYADFSHD |
| <i>A. niger</i> NRRL3135            | PTQGVGYaNE | LIARLTHSPV | HDDTSSNHTL | DSSPATFPLN | STLYADFSHD |
| <i>A. fumigatus</i> 13073           | PAQGIGFtNE | LIARLTRSPV | QDHTSTNsTL | vSNPATFPLN | ATMYVDFSHD |
| <i>A. fumigatus</i> 32722           | PAQGIGFtNE | LIARLTRSPV | QDHTSTNsTL | vSNPATFPLN | ATMYVDFSHD |
| <i>A. fumigatus</i> 58128           | PAQGIGFtNE | LIARLTRSPV | QDHTSTNsTL | vSNPATFPLN | ATMYVDFSHD |
| <i>A. fumigatus</i> 26906           | PAQGIGFtNE | LIARLTRSPV | QDHTSTNsTL | vSNPATFPLN | ATMYVDFSHD |
| <i>A. fumigatus</i> 32239           | PAQGIGFtNE | LIARLTNSPV | QDHTSTNsTL | DSDPATFPLN | ATIYVDFSHD |
| <i>E. nidulans</i>                  | PAQGIGFtNE | LIARLTQSPV | QDNTSTNHTL | DSNPATFPLD | rKLYADFSHD |
| <i>T. thermophilus</i>              | PAQGVGFvNE | LIARMTHSPV | QDYTTVNHTL | DSNPATFPLN | ATLYADFSHD |
| <i>M. thermophila</i>               | PTQGVGFvNE | LLARLAgvPV | RDgTSTNRTL | DGDPrTFPLG | rPLYADFSHD |
| Consensus                           | PAQGVGF-NE | LIARLTHSPV | QDHTSTNHTL | DSNPATFPLN | ATLYADFSHD |
| Consensus phytase                   | PAQGVGFANE | LIARLTRSPV | QDHTSTNHTL | DSNPATFPLN | ATLYADFSHD |

|                                     |            |            |            |            |            |
|-------------------------------------|------------|------------|------------|------------|------------|
|                                     | 351        |            |            | 400        |            |
| <i>A. terreus</i> 9A-1              | SNLVSIFWAL | GLYNGTAPLS | qTSVESVSQT | DGYAAAWTVP | FAARAYVEMM |
| <i>A. terreus</i> cbs               | SNLVSIFWAL | GLYNGTkPLS | qTTVEDITrT | DGYAAAWTVP | FAARAYIEMM |
| <i>A. niger</i> var. <i>awamori</i> | NGIISILFAL | GLYNGTkPLS | TTTVENITQT | DGFSSAWTVP | FASRLYVEMM |
| <i>A. niger</i> T213                | NGIISILFAL | GLYNGTkPLS | TTTVENITQT | DGFSSAWTVP | FASRLYVEMM |
| <i>A. niger</i> NRRL3135            | NGIISILFAL | GLYNGTkPLS | TTTVENITQT | DGFSSAWTVP | FASRLYVEMM |
| <i>A. fumigatus</i> 13073           | NSMVSIFFAL | GLYNGTEPLS | rTSVESaKEl | DGYSASWVVP | FGARAYFetM |
| <i>A. fumigatus</i> 32722           | NSMVSIFFAL | GLYNGTGPLS | rTSVESaKEl | DGYSASWVVP | FGARAYFetM |
| <i>A. fumigatus</i> 58128           | NSMVSIFFAL | GLYNGTEPLS | rTSVESaKEl | DGYSASWVVP | FGARAYFetM |
| <i>A. fumigatus</i> 26906           | NSMVSIFFAL | GLYNGTEPLS | rTSVESaKEl | DGYSASWVVP | FGARAYFetM |
| <i>A. fumigatus</i> 32239           | NGMIPIFFAM | GLYNGTEPLS | qTSeESTKES | NGYSASWAVP | FGARAYFetM |
| <i>E. nidulans</i>                  | NSMISIFFAM | GLYNGTQPLS | mDSVESIQEm | DGYAASWTVP | FGARAYFELM |
| <i>T. thermophilus</i>              | NTMTSIFaAL | GLYNGTAKLS | TTEIKSIEET | DGYSAAWTVP | FGGRAYIEMM |
| <i>M. thermophila</i>               | NDMMGVlGAL | GaYDGVPLD  | KTArrDpEEl | GGYAASWAVP | FAARiYVEKM |
| Consensus                           | NSMISIFFAL | GLYNGTAPLS | TTSVESIEET | DGYAASWTVP | FGARAYVEMM |
| Consensus phytase                   | NSMISIFFAL | GLYNGTAPLS | TTSVESIEET | DGYSASWTVP | FGARAYVEMM |

|                                     |             |            |            |            |            |
|-------------------------------------|-------------|------------|------------|------------|------------|
|                                     | 401         |            | 450        |            |            |
| <i>A. terreus</i> 9A-1              | QC.....     | .....RAEKE | PLVRVLVNDR | VMPLHGCPD  | KLGRCKrDAF |
| <i>A. terreus</i> cbs               | QC.....     | .....RAEKQ | PLVRVLVNDR | VMPLHGCAVD | NLGRCKrDDF |
| <i>A. niger</i> var. <i>awamori</i> | QC.....     | .....QAEQE | PLVRVLVNDR | VVPLHGCPID | aLGRCTrDSF |
| <i>A. niger</i> T213                | QC.....     | .....QAEQE | PLVRVLVNDR | VVPLHGCPID | aLGRCTrDSF |
| <i>A. niger</i> NRRL3135            | QC.....     | .....QAEQE | PLVRVLVNDR | VVPLHGCPVD | aLGRCTrDSF |
| <i>A. fumigatus</i> 13073           | QC.....     | .....KSEKE | PLVRALINDR | VVPLHGCDVD | KLGRCKLNDF |
| <i>A. fumigatus</i> 32722           | QC.....     | .....KSEKE | PLVRALINDR | VVPLHGCDVD | KLGRCKLNDF |
| <i>A. fumigatus</i> 58128           | QC.....     | .....KSEKE | SLVRALINDR | VVPLHGCDVD | KLGRCKLNDF |
| <i>A. fumigatus</i> 26906           | QC.....     | .....KSEKE | PLVRALINDR | VVPLHGCDVD | KLGRCKLNDF |
| <i>A. fumigatus</i> 32239           | QC.....     | .....KSEKE | PLVRALINDR | VVPLHGCAVD | KLGRCKLKDF |
| <i>E. nidulans</i>                  | QC.....     | .....E.KKE | PLVRVLVNDR | VVPLHGCAVD | KFGRCTLDWW |
| <i>T. thermophilus</i>              | QC.....     | .....DDSDE | PVVRVLVNDR | VVPLHGCEVD | SLGRCKrDDF |
| <i>M. thermophila</i>               | RCsgggggggg | ggegrQEKDE | eMVRVLVNDR | VMTLkGCGAD | ErGMCTLerF |
| Consensus                           | QC-----     | -----QAEKE | PLVRVLVNDR | VVPLHGCAVD | KLGRCKLDDF |
| Consensus phytase                   | QC.....     | .....QAEKE | PLVRVLVNDR | VVPLHGCAVD | KLGRCKRDDF |

Fig. 1c

|                            |      |     |     |      |
|----------------------------|------|-----|-----|------|
| Age                        | Mean | SD  | Min | Max  |
| Gender                     | 0.5  | 0.5 | 0   | 1    |
| Marital Status             | 0.6  | 0.5 | 0   | 1    |
| Education                  | 12.5 | 1.5 | 9   | 16   |
| Income                     | 1500 | 500 | 500 | 3000 |
| Health                     | 0.8  | 0.2 | 0   | 1    |
| Smoking                    | 0.3  | 0.5 | 0   | 1    |
| Alcohol                    | 0.2  | 0.4 | 0   | 1    |
| Exercise                   | 0.4  | 0.5 | 0   | 1    |
| Stress                     | 0.6  | 0.5 | 0   | 1    |
| Sleep                      | 0.7  | 0.3 | 0   | 1    |
| Work                       | 0.8  | 0.2 | 0   | 1    |
| Family                     | 0.9  | 0.1 | 0   | 1    |
| Friends                    | 0.7  | 0.4 | 0   | 1    |
| Community                  | 0.6  | 0.5 | 0   | 1    |
| Religion                   | 0.5  | 0.5 | 0   | 1    |
| Politics                   | 0.4  | 0.5 | 0   | 1    |
| Art                        | 0.3  | 0.5 | 0   | 1    |
| Music                      | 0.4  | 0.5 | 0   | 1    |
| Reading                    | 0.5  | 0.5 | 0   | 1    |
| Travel                     | 0.6  | 0.5 | 0   | 1    |
| Volunteering               | 0.7  | 0.4 | 0   | 1    |
| Charitable                 | 0.8  | 0.3 | 0   | 1    |
| Philanthropy               | 0.9  | 0.1 | 0   | 1    |
| Leadership                 | 0.6  | 0.5 | 0   | 1    |
| Teamwork                   | 0.7  | 0.4 | 0   | 1    |
| Communication              | 0.8  | 0.3 | 0   | 1    |
| Problem Solving            | 0.9  | 0.1 | 0   | 1    |
| Decision Making            | 0.7  | 0.4 | 0   | 1    |
| Conflict Resolution        | 0.6  | 0.5 | 0   | 1    |
| Emotional Stability        | 0.8  | 0.2 | 0   | 1    |
| Resilience                 | 0.7  | 0.3 | 0   | 1    |
| Optimism                   | 0.6  | 0.4 | 0   | 1    |
| Gratitude                  | 0.5  | 0.5 | 0   | 1    |
| Forgiveness                | 0.4  | 0.5 | 0   | 1    |
| Empathy                    | 0.7  | 0.3 | 0   | 1    |
| Compassion                 | 0.8  | 0.2 | 0   | 1    |
| Kindness                   | 0.9  | 0.1 | 0   | 1    |
| Generosity                 | 0.6  | 0.5 | 0   | 1    |
| Humility                   | 0.5  | 0.5 | 0   | 1    |
| Patience                   | 0.4  | 0.5 | 0   | 1    |
| Perseverance               | 0.7  | 0.3 | 0   | 1    |
| Self-Discipline            | 0.8  | 0.2 | 0   | 1    |
| Time Management            | 0.9  | 0.1 | 0   | 1    |
| Organization               | 0.6  | 0.5 | 0   | 1    |
| Productivity               | 0.7  | 0.4 | 0   | 1    |
| Focus                      | 0.8  | 0.3 | 0   | 1    |
| Attention                  | 0.9  | 0.1 | 0   | 1    |
| Memory                     | 0.6  | 0.5 | 0   | 1    |
| Learning                   | 0.7  | 0.4 | 0   | 1    |
| Adaptability               | 0.8  | 0.2 | 0   | 1    |
| Flexibility                | 0.9  | 0.1 | 0   | 1    |
| Openness                   | 0.6  | 0.5 | 0   | 1    |
| Conscientiousness          | 0.7  | 0.4 | 0   | 1    |
| Neuroticism                | 0.8  | 0.3 | 0   | 1    |
| Agreeableness              | 0.9  | 0.1 | 0   | 1    |
| Extraversion               | 0.6  | 0.5 | 0   | 1    |
| Introversion               | 0.5  | 0.5 | 0   | 1    |
| Social Skills              | 0.7  | 0.3 | 0   | 1    |
| Interpersonal Skills       | 0.8  | 0.2 | 0   | 1    |
| Leadership Skills          | 0.9  | 0.1 | 0   | 1    |
| Teamwork Skills            | 0.6  | 0.5 | 0   | 1    |
| Communication Skills       | 0.7  | 0.4 | 0   | 1    |
| Problem Solving Skills     | 0.8  | 0.3 | 0   | 1    |
| Decision Making Skills     | 0.9  | 0.1 | 0   | 1    |
| Conflict Resolution Skills | 0.6  | 0.5 | 0   | 1    |
| Emotional Stability Skills | 0.7  | 0.4 | 0   | 1    |
| Resilience Skills          | 0.8  | 0.2 | 0   | 1    |
| Optimism Skills            | 0.9  | 0.1 | 0   | 1    |
| Gratitude Skills           | 0.6  | 0.5 | 0   | 1    |
| Forgiveness Skills         | 0.7  | 0.4 | 0   | 1    |
| Empathy Skills             | 0.8  | 0.3 | 0   | 1    |
| Compassion Skills          | 0.9  | 0.1 | 0   | 1    |
| Kindness Skills            | 0.6  | 0.5 | 0   | 1    |
| Generosity Skills          | 0.7  | 0.4 | 0   | 1    |
| Humility Skills            | 0.8  | 0.2 | 0   | 1    |
| Patience Skills            | 0.9  | 0.1 | 0   | 1    |
| Perseverance Skills        | 0.6  | 0.5 | 0   | 1    |
| Self-Discipline Skills     | 0.7  | 0.4 | 0   | 1    |
| Time Management Skills     | 0.8  | 0.3 | 0   | 1    |
| Organization Skills        | 0.9  | 0.1 | 0   | 1    |
| Productivity Skills        | 0.6  | 0.5 | 0   | 1    |
| Focus Skills               | 0.7  | 0.4 | 0   | 1    |
| Attention Skills           | 0.8  | 0.2 | 0   | 1    |
| Memory Skills              | 0.9  | 0.1 | 0   | 1    |

| Variable                   | Mean | SD   | Min | Max  |
|----------------------------|------|------|-----|------|
| Age                        | 34.5 | 10.2 | 21  | 55   |
| Gender                     | 0.5  | 0.5  | 0   | 1    |
| Marital Status             | 0.6  | 0.5  | 0   | 1    |
| Education                  | 12.5 | 1.5  | 9   | 16   |
| Income                     | 1500 | 500  | 500 | 3000 |
| Health                     | 0.8  | 0.2  | 0   | 1    |
| Smoking                    | 0.3  | 0.5  | 0   | 1    |
| Alcohol                    | 0.2  | 0.4  | 0   | 1    |
| Exercise                   | 0.4  | 0.5  | 0   | 1    |
| Stress                     | 0.6  | 0.5  | 0   | 1    |
| Sleep                      | 0.7  | 0.3  | 0   | 1    |
| Work                       | 0.8  | 0.2  | 0   | 1    |
| Family                     | 0.9  | 0.1  | 0   | 1    |
| Friends                    | 0.7  | 0.4  | 0   | 1    |
| Community                  | 0.6  | 0.5  | 0   | 1    |
| Religion                   | 0.5  | 0.5  | 0   | 1    |
| Politics                   | 0.4  | 0.5  | 0   | 1    |
| Art                        | 0.3  | 0.5  | 0   | 1    |
| Music                      | 0.4  | 0.5  | 0   | 1    |
| Reading                    | 0.5  | 0.5  | 0   | 1    |
| Travel                     | 0.6  | 0.5  | 0   | 1    |
| Volunteering               | 0.7  | 0.4  | 0   | 1    |
| Charitable                 | 0.8  | 0.3  | 0   | 1    |
| Philanthropy               | 0.9  | 0.1  | 0   | 1    |
| Leadership                 | 0.6  | 0.5  | 0   | 1    |
| Teamwork                   | 0.7  | 0.4  | 0   | 1    |
| Communication              | 0.8  | 0.3  | 0   | 1    |
| Problem Solving            | 0.9  | 0.1  | 0   | 1    |
| Decision Making            | 0.7  | 0.4  | 0   | 1    |
| Conflict Resolution        | 0.6  | 0.5  | 0   | 1    |
| Emotional Stability        | 0.8  | 0.2  | 0   | 1    |
| Resilience                 | 0.7  | 0.3  | 0   | 1    |
| Optimism                   | 0.6  | 0.4  | 0   | 1    |
| Gratitude                  | 0.5  | 0.5  | 0   | 1    |
| Forgiveness                | 0.4  | 0.5  | 0   | 1    |
| Empathy                    | 0.7  | 0.3  | 0   | 1    |
| Compassion                 | 0.8  | 0.2  | 0   | 1    |
| Kindness                   | 0.9  | 0.1  | 0   | 1    |
| Generosity                 | 0.6  | 0.5  | 0   | 1    |
| Humility                   | 0.5  | 0.5  | 0   | 1    |
| Patience                   | 0.4  | 0.5  | 0   | 1    |
| Perseverance               | 0.7  | 0.3  | 0   | 1    |
| Self-Discipline            | 0.8  | 0.2  | 0   | 1    |
| Time Management            | 0.9  | 0.1  | 0   | 1    |
| Organization               | 0.6  | 0.5  | 0   | 1    |
| Productivity               | 0.7  | 0.4  | 0   | 1    |
| Focus                      | 0.8  | 0.3  | 0   | 1    |
| Attention                  | 0.9  | 0.1  | 0   | 1    |
| Memory                     | 0.6  | 0.5  | 0   | 1    |
| Learning                   | 0.7  | 0.4  | 0   | 1    |
| Adaptability               | 0.8  | 0.2  | 0   | 1    |
| Flexibility                | 0.9  | 0.1  | 0   | 1    |
| Openness                   | 0.6  | 0.5  | 0   | 1    |
| Conscientiousness          | 0.7  | 0.4  | 0   | 1    |
| Neuroticism                | 0.8  | 0.3  | 0   | 1    |
| Agreeableness              | 0.9  | 0.1  | 0   | 1    |
| Extraversion               | 0.6  | 0.5  | 0   | 1    |
| Introversion               | 0.5  | 0.5  | 0   | 1    |
| Social Skills              | 0.7  | 0.3  | 0   | 1    |
| Interpersonal Skills       | 0.8  | 0.2  | 0   | 1    |
| Leadership Skills          | 0.9  | 0.1  | 0   | 1    |
| Teamwork Skills            | 0.6  | 0.5  | 0   | 1    |
| Communication Skills       | 0.7  | 0.4  | 0   | 1    |
| Problem Solving Skills     | 0.8  | 0.3  | 0   | 1    |
| Decision Making Skills     | 0.9  | 0.1  | 0   | 1    |
| Conflict Resolution Skills | 0.6  | 0.5  | 0   | 1    |
| Emotional Stability Skills | 0.7  | 0.4  | 0   | 1    |
| Resilience Skills          | 0.8  | 0.2  | 0   | 1    |
| Optimism Skills            | 0.9  | 0.1  | 0   | 1    |
| Gratitude Skills           | 0.6  | 0.5  | 0   | 1    |
| Forgiveness Skills         | 0.7  | 0.4  | 0   | 1    |
| Empathy Skills             | 0.8  | 0.3  | 0   | 1    |
| Compassion Skills          | 0.9  | 0.1  | 0   | 1    |
| Kindness Skills            | 0.6  | 0.5  | 0   | 1    |
| Generosity Skills          | 0.7  | 0.4  | 0   | 1    |
| Humility Skills            | 0.8  | 0.2  | 0   | 1    |
| Patience Skills            | 0.9  | 0.1  | 0   | 1    |
| Perseverance Skills        | 0.6  | 0.5  | 0   | 1    |
| Self-Discipline Skills     | 0.7  | 0.4  | 0   | 1    |
| Time Management Skills     | 0.8  | 0.3  | 0   | 1    |
| Organization Skills        | 0.9  | 0.1  | 0   | 1    |
| Productivity Skills        | 0.6  | 0.5  | 0   | 1    |
| Focus Skills               | 0.7  | 0.4  | 0   | 1    |
| Attention Skills           | 0.8  | 0.2  | 0   | 1    |
| Memory Skills              | 0.9  | 0.1  | 0   | 1    |
|                            |      |      |     |      |

CP-1  
 Eco RI M G V F V V L L S I A T L F G S T  
**TATATGAATTCATGGGCGTGTTCGTGCTACTGTCCATTGCCACCTTGTTCGGTTCCA**  
 1 -----+-----+-----+-----+-----+-----+ 60  
 ATATACTTAAGTACCCGCACAAGCAGCAGATGACAGGTAACGGTGGAACAAGCCAAGGT  
  
 S G T A L G P R G N S H S C D T V D G G  
**CATCCGGTACCGCCTTGGGTCCTCGTGGAATTCTCACTCTTGTGACACTGTTGACGGTG**  
 61 -----+-----+-----+-----+-----+ 120  
 GTAGGCCATGGCGGAACCCAGGAGCACCATTAAAGAGTGAGAACACTGTGACAACCTGCCAC  
 CP-2  
 CP-3  
 Y Q C F P E I S H L W G Q Y S P Y F S L  
 GTTACCAATGTTTCCAGAAATTTCTCACTTGTGGGGTCAATACTCTCCATACTTCTCTT  
 121 -----+-----+-----+-----+-----+ 180  
 CAATGGTTACAAAGGGTCTTTAAAGAGTGAACACCCAGTTATGAGAGGTATGAAGAGAA  
  
 E D E S A I S P D V P D D C R V T F V Q  
**TGGAAGACGAATCTGCTATTTCTCCAGACGTTCCAGACGACTGTAGAGTTACTTTCGTTT**  
 181 -----+-----+-----+-----+-----+ 240  
 ACCTTCTGCTTAGACGATAAAGAGGTCTGCAAGGTCTGCTGACATCTCAATGAAAGCAAG  
 CP-4  
 CP-5  
 V L S R H G A R Y P T S S K S K A Y S A  
 AAGTTTTGTCTAGACACGGTGCTAGATACCCAACTTCTTCTAAGTCTAAGGCTTACTCTG  
 241 -----+-----+-----+-----+-----+ 300  
 TTCAAACAGATCTGTGCCACGATCTATGGGTTGAAGAAGATTGAGATTCCGAATGAGAC  
  
 L I E A I Q K N A T A F K G K Y A F L K  
**CTTTGATTGAAGCTATTCAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGA**  
 301 -----+-----+-----+-----+-----+ 360  
 GAAACTAACTTCGATAAGTTTTCTTGGCATGACGAAAGTTCCCATTCATGCGAAAGAAGT  
 CP-6  
 CP-7  
 T Y N Y T L G A D D L T P F G E N Q M V  
 AGACTTACAACCTACACTTTGGGTGCTGACGACTTGACTCCATTGCGTGAAAACCAAATGG  
 361 -----+-----+-----+-----+-----+ 420  
 TCTGAATGTTGATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACC  
  
 N S G I K F Y R R Y K A L A R K I V P F  
**TTAACTCTGGTATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCAT**  
 421 -----+-----+-----+-----+-----+ 480  
 AATTGAGACCATAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTA  
 CP-8  
 CP-9  
 I R A S G S D R V I A S A E K F I E G F  
 TCATTAGAGCTTCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCAATTGAAGGTT  
 481 -----+-----+-----+-----+-----+ 540  
 AGTAATCTCGAAGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACCTTCCAA  
  
 Q S A K L A D P G S Q P H Q A S P V I D  
**TCCAATCTGCTAAGTTGGCTGACCCAGGTTCTCAACCACACCAAGCTTCTCCAGTTATTG**  
 541 -----+-----+-----+-----+-----+ 600  
 AGGTTAGACGATTCAACCGACTGGGTCCAAGAGTTGGTGTGGTTTCAAGAGGTCAATAAC

Fig. 2a

**Fig. 2b**

| Variable              | Mean | SD   | Min | Max  |
|-----------------------|------|------|-----|------|
| Age                   | 34.5 | 10.2 | 18  | 65   |
| Gender                | 0.5  | 0.5  | 0   | 1    |
| Marital status        | 0.6  | 0.5  | 0   | 1    |
| Education             | 12.5 | 1.5  | 9   | 16   |
| Income                | 1500 | 500  | 500 | 3000 |
| Health status         | 0.7  | 0.4  | 0   | 1    |
| Smoking status        | 0.3  | 0.5  | 0   | 1    |
| Alcohol consumption   | 0.2  | 0.4  | 0   | 1    |
| Exercise frequency    | 0.5  | 0.5  | 0   | 1    |
| Stress level          | 0.6  | 0.5  | 0   | 1    |
| Sleep quality         | 0.7  | 0.4  | 0   | 1    |
| Work satisfaction     | 0.6  | 0.5  | 0   | 1    |
| Life satisfaction     | 0.7  | 0.4  | 0   | 1    |
| Depression score      | 10.5 | 5.0  | 0   | 30   |
| Anxiety score         | 12.0 | 6.0  | 0   | 30   |
| Quality of life score | 75.0 | 10.0 | 50  | 100  |

| Variable                             | Mean | SD   | Min | Max  |
|--------------------------------------|------|------|-----|------|
| Age                                  | 34.5 | 10.2 | 18  | 65   |
| Gender                               | 0.5  | 0.5  | 0   | 1    |
| Marital status                       | 0.6  | 0.5  | 0   | 1    |
| Education                            | 12.5 | 1.5  | 9   | 16   |
| Income                               | 1500 | 500  | 500 | 3000 |
| Health status                        | 0.7  | 0.4  | 0   | 1    |
| Smoking status                       | 0.3  | 0.5  | 0   | 1    |
| Alcohol consumption                  | 0.2  | 0.4  | 0   | 1    |
| Exercise frequency                   | 0.5  | 0.5  | 0   | 1    |
| Stress level                         | 0.6  | 0.4  | 0   | 1    |
| Sleep quality                        | 0.7  | 0.3  | 0   | 1    |
| Work satisfaction                    | 0.6  | 0.4  | 0   | 1    |
| Life satisfaction                    | 0.7  | 0.3  | 0   | 1    |
| Depression score                     | 0.3  | 0.4  | 0   | 1    |
| Anxiety score                        | 0.2  | 0.3  | 0   | 1    |
| Quality of life score                | 0.6  | 0.4  | 0   | 1    |
| Healthcare utilization               | 0.4  | 0.5  | 0   | 1    |
| Health insurance status              | 0.8  | 0.4  | 0   | 1    |
| Chronic disease status               | 0.1  | 0.3  | 0   | 1    |
| Family size                          | 2.5  | 1.0  | 1   | 5    |
| Home ownership                       | 0.7  | 0.4  | 0   | 1    |
| Commute time                         | 15   | 10   | 5   | 45   |
| Neighborhood safety                  | 0.8  | 0.3  | 0   | 1    |
| Access to healthcare                 | 0.9  | 0.2  | 0   | 1    |
| Healthcare costs                     | 100  | 50   | 50  | 200  |
| Healthcare quality                   | 0.7  | 0.3  | 0   | 1    |
| Healthcare accessibility             | 0.8  | 0.3  | 0   | 1    |
| Healthcare affordability             | 0.6  | 0.4  | 0   | 1    |
| Healthcare satisfaction              | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization frequency     | 0.5  | 0.5  | 0   | 1    |
| Healthcare utilization duration      | 10   | 5    | 5   | 20   |
| Healthcare utilization cost          | 50   | 25   | 25  | 100  |
| Healthcare utilization quality       | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization accessibility | 0.8  | 0.3  | 0   | 1    |
| Healthcare utilization affordability | 0.6  | 0.4  | 0   | 1    |
| Healthcare utilization satisfaction  | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization frequency     | 0.5  | 0.5  | 0   | 1    |
| Healthcare utilization duration      | 10   | 5    | 5   | 20   |
| Healthcare utilization cost          | 50   | 25   | 25  | 100  |
| Healthcare utilization quality       | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization accessibility | 0.8  | 0.3  | 0   | 1    |
| Healthcare utilization affordability | 0.6  | 0.4  | 0   | 1    |
| Healthcare utilization satisfaction  | 0.7  | 0.3  | 0   | 1    |

| Variable                             | Mean | SD   | Min | Max  |
|--------------------------------------|------|------|-----|------|
| Age                                  | 34.5 | 10.2 | 18  | 65   |
| Gender                               | 0.5  | 0.5  | 0   | 1    |
| Marital status                       | 0.6  | 0.5  | 0   | 1    |
| Education                            | 12.5 | 1.5  | 9   | 16   |
| Income                               | 1500 | 500  | 500 | 3000 |
| Health status                        | 0.7  | 0.4  | 0   | 1    |
| Smoking status                       | 0.3  | 0.5  | 0   | 1    |
| Alcohol consumption                  | 0.2  | 0.4  | 0   | 1    |
| Exercise frequency                   | 0.5  | 0.5  | 0   | 1    |
| Stress level                         | 0.6  | 0.4  | 0   | 1    |
| Sleep quality                        | 0.7  | 0.3  | 0   | 1    |
| Work satisfaction                    | 0.6  | 0.4  | 0   | 1    |
| Life satisfaction                    | 0.7  | 0.3  | 0   | 1    |
| Depression score                     | 0.3  | 0.4  | 0   | 1    |
| Anxiety score                        | 0.2  | 0.3  | 0   | 1    |
| Quality of life score                | 0.6  | 0.4  | 0   | 1    |
| Healthcare utilization               | 0.4  | 0.5  | 0   | 1    |
| Health insurance status              | 0.8  | 0.4  | 0   | 1    |
| Chronic disease status               | 0.1  | 0.3  | 0   | 1    |
| Family size                          | 2.5  | 1.0  | 1   | 5    |
| Home ownership                       | 0.7  | 0.4  | 0   | 1    |
| Commute time                         | 15   | 10   | 5   | 45   |
| Neighborhood safety                  | 0.8  | 0.3  | 0   | 1    |
| Access to healthcare                 | 0.9  | 0.2  | 0   | 1    |
| Healthcare costs                     | 100  | 50   | 50  | 200  |
| Healthcare quality                   | 0.7  | 0.3  | 0   | 1    |
| Healthcare accessibility             | 0.8  | 0.3  | 0   | 1    |
| Healthcare affordability             | 0.6  | 0.4  | 0   | 1    |
| Healthcare satisfaction              | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization frequency     | 0.5  | 0.5  | 0   | 1    |
| Healthcare utilization duration      | 10   | 5    | 5   | 20   |
| Healthcare utilization cost          | 50   | 25   | 25  | 100  |
| Healthcare utilization quality       | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization accessibility | 0.8  | 0.3  | 0   | 1    |
| Healthcare utilization affordability | 0.6  | 0.4  | 0   | 1    |
| Healthcare utilization satisfaction  | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization frequency     | 0.5  | 0.5  | 0   | 1    |
| Healthcare utilization duration      | 10   | 5    | 5   | 20   |
| Healthcare utilization cost          | 50   | 25   | 25  | 100  |
| Healthcare utilization quality       | 0.7  | 0.3  | 0   | 1    |
| Healthcare utilization accessibility | 0.8  | 0.3  | 0   | 1    |
| Healthcare utilization affordability | 0.6  | 0.4  | 0   | 1    |
| Healthcare utilization satisfaction  | 0.7  | 0.3  | 0   | 1    |

|                             |            |            |            |            |            |
|-----------------------------|------------|------------|------------|------------|------------|
|                             | 1          |            |            |            | 50         |
| <i>P. involutus</i> (phyA1) | SvP.KnTAPt | FPIPeseQrn | WSPYSPYFPL | AeYkAPPAGC | QInQVNIIQR |
| <i>P. involutus</i> (phyA2) | SvP.RniAPK | FSIPeseQrn | WSPYSPYFPL | AeYkAPPAGC | EInQVNIIQR |
| <i>T. pubescens</i>         | hiPlRdTSAC | LdVTrDvQqs | WSmYSPYFPa | AtYvAPPASC | QInQVHIIQR |
| <i>A. pediades</i>          | GgvvQaTfvQ | pfFPpQiQds | WAAyTPYYPV | qaYtPPPkDC | KItQVNIIQR |
| <i>P. lycii</i>             | StQfsfvAAQ | LPIPaQntsn | WGPYdPFFPV | EpYaAPPEGC | tVtQVNLIQR |

**Basidio**                      **S-P-R-TAAQ**   **LPIP-Q-Q--**   **WSPYSPYFPV**   **A-Y-APPAGC**   **QI-QVNIIQR**

|                             |            |            |            |            |            |
|-----------------------------|------------|------------|------------|------------|------------|
|                             | 51         |            |            |            | 100        |
| <i>P. involutus</i> (phyA1) | HGARFPTSGA | TTRIKAGLTK | LQGvqnftDA | KFNFIksfky | dLGnsDLVPP |
| <i>P. involutus</i> (phyA2) | HGARFPTSGA | ATRIKAGLSK | LQSVqnftDP | KFDFIksfTY | dLGtsDLVPP |
| <i>T. pubescens</i>         | HGARFPTSGA | AKRIQTAVAK | LKAAsnyTDP | lLAFvtNyTY | sLGqDsLLeL |
| <i>A. pediades</i>          | HGARFPTSGA | GTRIQAaVvK | LQSAktyTDP | RLDFLTnyTY | tLGhDDLVPF |
| <i>P. lycii</i>             | HGARWPTSGA | rSRqvAAVAK | IQmArpftDP | KYEFLnDfvY | kFGvADLLPF |

**Basidio**                      **HGARFPTSGA**   **ATRIQAaVAK**   **LQSA---TDP**   **KLDFL-N-TY**   **-LG-DDLVPP**

|                             |            |            |            |            |            |
|-----------------------------|------------|------------|------------|------------|------------|
|                             | 101        |            |            |            | 150        |
| <i>P. involutus</i> (phyA1) | GAAQSfDAGQ | EAFARYSkLV | SkNNLPFIRA | dGSDRVVDSA | TNWTAGFAsA |
| <i>P. involutus</i> (phyA2) | GAAQSfDAGl | EvFARYSkLV | SsDNLPFIRS | dGSDRVVDTA | TNWTAGFAsA |
| <i>T. pubescens</i>         | GAtQSSEAGQ | EAFTRYSSLV | SaDELpFVRA | SGSDRVVATA | nNWTAGFALA |
| <i>A. pediades</i>          | GAlQSSQAGE | ETFqRYSfLV | SkENLPFVRA | SSSNRVVDSA | TNWTGFSaA  |
| <i>P. lycii</i>             | GAnQShQTGt | DmYTRYStLf | egGDVPFVRA | AGdQRVVDSS | TNWTAGFGdA |

**Basidio**                      **GA-QSSQAGQ**   **EAFTRYs-LV**   **S-DNLpFVRA**   **SGSDRVVDSA**   **TNWTAGFA-A**

|                             |            |            |            |            |            |
|-----------------------------|------------|------------|------------|------------|------------|
|                             | 151        |            |            |            | 200        |
| <i>P. involutus</i> (phyA1) | ShNTVqPkLn | LILPQtGNDT | LEDNMCPaAG | DSDPQvNaWL | AVafPSITAR |
| <i>P. involutus</i> (phyA2) | SrNAiqPkLd | LILPQtGNDT | LEDNMCPaAG | ESDPQvDaWL | AsafPSVTAQ |
| <i>T. pubescens</i>         | SsNSitPvLs | VIIEEaGNDT | LDDNMCPaAG | DSDPQvNqWL | AqFAPPMTAR |
| <i>A. pediades</i>          | ShHvlnPiLf | VILSEsLNDT | LDDaMCPnAG | sSDPQtGiWt | SIYGTPIAnR |
| <i>P. lycii</i>             | SgETvlPtLq | VVLqEeGNcT | LcNNMCPnEv | DGDest.tWL | GVFAPnITAR |

**Basidio**                      **S-NT--P-L-**   **VILSE-GNDT**   **LDDNMCP-AG**   **DSDPQ-N-WL**   **AVFAPPITAR**

|                             |            |            |            |            |            |
|-----------------------------|------------|------------|------------|------------|------------|
|                             | 201        |            |            |            | 250        |
| <i>P. involutus</i> (phyA1) | LNAAAPSVNL | TDtDAfNLvs | LCAFlTVSke | kkSdFctLFE | giPGsFeAFa |
| <i>P. involutus</i> (phyA2) | LNAAAPGANL | TDaDAfNLvs | LCPFmTVSke | qkSdFctLFE | giPGsFeAFa |
| <i>T. pubescens</i>         | LNAGAPGANL | TDtDTyNLlt | LCPFETVate | rrSeFCDIYE | elQAE.dAFa |
| <i>A. pediades</i>          | LNqqAPGANI | TAAvdsNLip | LCAFETivke | tpSpFCNLf. | .tPEEFaqFe |
| <i>P. lycii</i>             | LNAAAPSANL | SDsDAltLmd | MCPFDTLSSg | naSpFCDLf. | .tAEEYvSYe |

**Basidio**                      **LNAAAPGANL**   **TD-DA-NL--**   **LCPFETVS-E**   **--S-FCDLFE**   **--PEEF-AF-**

Fig. 3a

251 300

*P. involutus* (phyA1) YgGDLDKfYG TGYGQeLGPV QGVGYVNElI ARLTnsAVRD NTQTNRtLDA

*P. involutus* (phyA2) YaGDLDKfYG TGYGQALGPV QGVGYINELL ARLTnsAVnD NTQTNRtLDA

*T. pubescens* YnADLDKfYG TGYGQPLGPV QGVGYINELI ARLTaQnVsD HTQTNsTLDS

*A. pediades* YfGDLDKfYG TGYGQPLGPV QGVGYINELL ARLTemPVrD NTQTNRtLDS

*P. lycii* YyyDLdKYyG TGpGNALGPV QGVGYVNELL ARLTgQAVRD ETQTNRtLDS

**Basidio** **Y-GDLDKfYG TGYGQPLGPV QGVGYINELL ARLT-QAVRD NTQTNRtLDS**

301 350

*P. involutus* (phyA1) SPvTFPLNKT FYADFSHDNl MVAVFSAAGL FrQPAPLsTS vPNPwRTWrT

*P. involutus* (phyA2) APdTFPLNKT MYADFSHDNl MVAVFSAAGL FrQSAPLsTS tPDPNRTWLT

*T. pubescens* SPeTFPLNRT LYADFSHDNQ MVAIFSAAGL FNQSAPLDPT tPDPaRTFLv

*A. pediades* SPtTFPLDRS IYADLSHDNQ MIAIFSAAGL FNQSSPLDPS fPNPKRTWVT

*P. lycii* dPaTFPLNRT FYADFSHDNt MVPIFAALGL FNATa.LDPl kPDENRlWVD

**Basidio** **SP-TFPLNRT FYADFSHDNQ MVAIFSAAGL FNQSAPLDPS -PDPNRTWVT**

351 400

*P. involutus* (phyA1) SsLVPFSGRM VVERLsC..f GT..... .tkv RVLVQDqVQP

*P. involutus* (phyA2) SsVVPFSARM aVERLsC..a GT..... .tkv RVLVQDqVQP

*T. pubescens* kKIVPFSGRM VVERLdC..g GA..... .qsV RLLVNDAVQP

*A. pediades* SRLtPFSGRM VtERLlCqrd GTgsgggsri mrngnvqtfv RILVNDAVQP

*P. lycii* SKLVPFSGHM tVEKLsC... .sgkeaV RVLVNDAVQP

**Basidio** **SKLVPFSGRM VVERL-C--- GT----- -V RVLVNDAVQP**

401 441

*P. involutus* (phyA1) LEFCGGDrNG lCTLAkfVES QtFARsDGaG DFekCFATSa ~

*P. involutus* (phyA2) LEFCGGDqDG lCALDkfVES QaYARsGGaG DFekCLATTv ~

*T. pubescens* LAFCGADtsG vCTLDafVES QaYARNDGEG DFekCFAT~~ ~

*A. pediades* LKFCGGDmDS lCTLEafVES QkYAREdGQG DFekCFD~~~ ~

*P. lycii* LEFCGG.vDG vCeLsAfVES QtYARENGQG DfAKCgfvPs e

**Basidio** **LEFCGGD-DG -CTLDafVES Q-YAREdGQG DFekCFATP- -**

Fig. 3b

|                                     |            |             |                                  |
|-------------------------------------|------------|-------------|----------------------------------|
|                                     | 1          |             | 50                               |
| <i>A. terreus</i> 9a1               | KhSDCNSVDh | GYQCfPELSH  | kWGLYAPYFS LqDESPFP1D VPdDCHITFV |
| <i>A. terreus</i> cbs               | NhSDCtSVDr | GYQCfPELSH  | kWGLYAPYFS LqDESPFP1D VPdDCHITFV |
| <i>A. niger</i> var. <i>awamori</i> | NqSTCDTVdQ | GYQCfSEtSH  | LWGQYAPFFS LANESAISPD VPAGCRVTFa |
| <i>A. niger</i> NRRL3135            | NqSSCDTVdQ | GYQCfSEtSH  | LWGQYAPFFS LANESvISPE VPAGCRVTFa |
| <i>A. fumigatus</i> 13073           | GSKSCDTVD1 | GYQCSPatSH  | LWGQYSPFFS LEDELsvSSK LpKDCRITLV |
| <i>A. fumigatus</i> 32722           | GSKSCDTVD1 | GYQCSPatSH  | LWGQYSPFFS LEDELsvSSK LpKDCRITLV |
| <i>A. fumigatus</i> 58128           | GSKSCDTVD1 | GYQCSPatSH  | LWGQYSPFFS LEDELsvSSK LpKDCRITLV |
| <i>A. fumigatus</i> 26906           | GSKSCDTVD1 | GYQCSPatSH  | LWGQYSPFFS LEDELsvSSK LpKDCRITLV |
| <i>A. fumigatus</i> 32239           | GSKACDTVEL | GYQCSPGtSH  | LWGQYSPFFS LEDELsvSSD LpKDCRVTFV |
| <i>E. nidulans</i>                  | QNHSCNTaDG | GYQCfPNVSH  | VWGQYSPYFS IEQESAISeD VPhGCeVTFV |
| <i>T. thermophilus</i>              | DSHSCNTVEG | GYQCrPEISH  | sWGQYSPFFS LADQSEISPD VPqNCKITFV |
| <i>T. lanuginosus</i>               | ~~~~~      | ~~~~~nvDIAR | hWGQYSPFFS LAEvSEISPA VPkGCRVeFV |
| <i>M. thermophila</i>               | ESRPCDTpD1 | GFQCgTAISH  | FWGQYSPYFS VPSElDaS.. IPdDCeVTFa |
| Basidio                             | xSxPxrxTAA | qLPipxQxqx  | xWSPYSPYFP VAXyxA.... pPaGcQIxqV |

|           |            |            |            |            |            |
|-----------|------------|------------|------------|------------|------------|
| Consensus | NSHSCDTVDG | GYQC-PEISH | LWGQYSPFFS | LADESAISPD | VP-GCRVTFV |
| Fcp10     | NSHSCDTVDG | GYQCfPEISH | LWGQYSPFFS | LADESAISPD | VPKGCRVTFV |

|                                     |            |            |                                   |
|-------------------------------------|------------|------------|-----------------------------------|
|                                     | 51         |            | 100                               |
| <i>A. terreus</i> 9a1               | QVLARHGARS | PTHSKTKaYA | AtIaAIQKSA TaFpGKYAFL QSYNYSLDSE  |
| <i>A. terreus</i> cbs               | QVLARHGARS | PTdSKTKaYA | AtIaAIQKNA TaLpGKYAFL KSYNYSMGSE  |
| <i>A. niger</i> var. <i>awamori</i> | QVLSRHGARY | PTeSKGKKYS | ALIEeIQQNv TtFDGKYAFL KTYNYSLGAD  |
| <i>A. niger</i> NRRL3135            | QVLSRHGARY | PTdSKGKKYS | ALIEeIQQNA TtFDGKYAFL KTYNYSLGAD  |
| <i>A. fumigatus</i> 13073           | QVLSRHGARY | PTSSKSKKYk | kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD  |
| <i>A. fumigatus</i> 32722           | QVLSRHGARY | PTSSKSKKYk | kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD  |
| <i>A. fumigatus</i> 58128           | QVLSRHGARY | PTSSKSKKYk | kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD  |
| <i>A. fumigatus</i> 26906           | QVLSRHGARY | PTSSKSKKYk | kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD  |
| <i>A. fumigatus</i> 32239           | QVLSRHGARY | PTASKSKKYk | kLVtAIQKNA TeFKGKFAFL ETYNYTLGAD  |
| <i>E. nidulans</i>                  | QVLSRHGARY | PTeSKSKaYS | GLIEAIQKNA TsFwGQYAFI ESYNITLGAD  |
| <i>T. thermophilus</i>              | QLLSRHGARY | PTSSKTELYS | qLIrIQKtA TaYKGyYAFI KdYrYqLGAN   |
| <i>T. lanuginosus</i>               | QVLSRHGARY | PTAhKSEvYA | ELLqrIQDtA TeFKGDFAFL RdYaYhLGAD  |
| <i>M. thermophila</i>               | QVLSRHGARA | PTlkRAAsYv | DLIdrIHhGA isYgPgYEFL RTYDNYTLGAD |
| Basidio                             | NIIqRHGARF | PTSGaAtRiq | AaVakLQsax xxtDPKLDLFL xnxtYxLGxD |

|           |            |            |            |            |            |
|-----------|------------|------------|------------|------------|------------|
| Consensus | QVLSRHGARY | PTSSKSKKYS | ALI-AIQKNA | T-FKGKYAFL | KTYNYTLGAD |
| Fcp10     | QVLSRHGARY | PTSSKSKKYS | ALIEAIQKNA | TAFKGKYAFL | KTYNYTLGAD |

|                                     |            |            |                                  |
|-------------------------------------|------------|------------|----------------------------------|
|                                     | 101        |            | 150                              |
| <i>A. terreus</i> 9a1               | ELTPFGrNQL | rDlGaQFYeR | YNAL.TRhIn PFVRATDAsR VhESAeKFVE |
| <i>A. terreus</i> cbs               | NLTPFGrNQL | qDlGaQFYRR | YDTL.TRhIn PFVRAADSSr VhESAeKFVE |
| <i>A. niger</i> var. <i>awamori</i> | DLTPFGEQEL | VNSGIKFYQR | YESL.TRnII PFIRSSGSsR VIASGEKFIE |
| <i>A. niger</i> NRRL3135            | DLTPFGEQEL | VNSGIKFYQR | YESL.TRnIV PFIRSSGSsR VIASGKKFIE |
| <i>A. fumigatus</i> 13073           | DLTPFGEQQL | VNSGIKFYQR | YKAL.ARsVV PFIRASGSDR VIASGEKFIE |
| <i>A. fumigatus</i> 32722           | DLTPFGEQQL | VNSGIKFYQR | YKAL.ARsVV PFIRASGSDR VIASGEKFIE |
| <i>A. fumigatus</i> 58128           | DLTPFGEQQL | VNSGIKFYQR | YKAL.ARsVV PFIRASGSDR VIASGEKFIE |
| <i>A. fumigatus</i> 26906           | DLTAfGEQQL | VNSGIKFYQR | YKAL.ARsVV PFIRASGSDR VIASGEKFIE |
| <i>A. fumigatus</i> 32239           | DLTPFGEQQM | VNSGIKFYQK | YKAL.AgsVV PFIRSSGSsR VIASGEKFIE |
| <i>E. nidulans</i>                  | DLTiFGENQM | VDSGaKFYRR | YKnL.ARKnt PFIRASGSDR VVASAEKFIE |
| <i>T. thermophilus</i>              | DLTPFGENQM | IQLGIKFYnH | YKSL.ARnaV PFVRCSGSDR VIASGrIFIE |
| <i>T. lanuginosus</i>               | NLTrFGEEQM | MESGrQFYHR | YREq.AReIV PFVRAAGSAR VIASAEfFnr |
| <i>M. thermophila</i>               | ELTrtGOQQM | VNSGIKFYRR | YRAL.ARKsI PFVRTAGqDR VVhSAENftQ |
| Basidio                             | DLvPFGAxQs | sQAGqEaFtR | YsxLvSxdnL PFVRASGSDR VVDSAtNwtA |

|           |            |            |            |            |            |
|-----------|------------|------------|------------|------------|------------|
| Consensus | DLTPFGEQQM | VNSGIKFYRR | YKAL-AR-IV | PFVRASGSDR | VIASAEKFIE |
| Fcp10     | DLTPFGEQQM | VNSGIKFYRR | YKAL.ARKIV | PFVRASGSDR | VIASAEKFIE |

Fig. 4a

|                                     |            |            |                                   |
|-------------------------------------|------------|------------|-----------------------------------|
|                                     | 151        |            | 200                               |
| <i>A. terreus</i> 9a1               | GFQTARqDDh | hAnphQPSPr | VDVaIPEGsA YNNTLEHSLC TAFES...St  |
| <i>A. terreus</i> cbs               | GFQNARqGDP | hAnphQPSPr | VDVVIPEGtA YNNTLEHSIC TAFEa...St  |
| <i>A. niger</i> var. <i>awamori</i> | GFQSTKLkDP | rAqpgQSSPk | IDVVISEAsS sNNTLDpGtC TvFed...SE  |
| <i>A. niger</i> NRRL3135            | GFQSTKLkDP | rAqpgQSSPk | IDVVISEAsS sNNTLDpGtC TvFed...SE  |
| <i>A. fumigatus</i> 13073           | GFQqAKLADP | gAt.nRAAPa | ISVIIPESeT FNNTLDHGVC TkFEa...SQ  |
| <i>A. fumigatus</i> 32722           | GFQqAKLADP | gAt.nRAAPa | ISVIIPESeT FNNTLDHGVC TkFEa...SQ  |
| <i>A. fumigatus</i> 58128           | GFQqAKLADP | gAt.nRAAPa | ISVIIPESeT FNNTLDHGVC TkFEa...SQ  |
| <i>A. fumigatus</i> 26906           | GFQqAKLADP | gAt.nRAAPa | ISVIIPESeT FNNTLDHGVC TkFEa...SQ  |
| <i>A. fumigatus</i> 32239           | GFQqANVADP | gAt.nRAAPV | ISVIIPESeT YNNTLDHSVC TnFEa...SE  |
| <i>E. nidulans</i>                  | GFRkAQLhDh | g.s.gQATPV | VNVIIEIdG FNNTLDHStC vSFEn...dE   |
| <i>T. thermophilus</i>              | GFQSAKVLDP | hSdKhDAPpt | INVIIeEGpS YNNTLDtGsC PvFed...Ss  |
| <i>T. lanuginosus</i>               | GFQdAKdrDP | rSnkdQAePV | INVIISEEtG sNNTLDgltC PAaEe...Ap  |
| <i>M. thermophila</i>               | GFHSALLADR | gStvrPTlPy | dmVVIPETaG aNNTLHNDLC TAFEegPySt  |
| Basidio                             | GFaxA..... | ..sxntxxPx | LxVILSExg. .NDTLDDNMC .....PxAG   |
| Consensus                           | GFQSAKLADP | -A---QASPV | INVIIPEG-G YNNTLDHGLC TAFE--P-SE  |
| Fcp10                               | GFQSAKLADP | GANPHQASPV | INVIIPEGAG YNNTLDHGLC TAFE...SE   |
|                                     | 201        |            | 250                               |
| <i>A. terreus</i> 9a1               | VGDDaVANFT | AVFAPAIaQr | LEAdLPGVQL StDDVVNLMA MCPFETVSLT  |
| <i>A. terreus</i> cbs               | VGDAaADNFT | AVFAPAIaKr | LEAdLPGVQL SADDVVNLMA MCPFETVSLT  |
| <i>A. niger</i> var. <i>awamori</i> | LADtVEANFT | AtFAPSIRqR | LEndLSGVtL TDtEVtyLMD MCSFDTISTs  |
| <i>A. niger</i> NRRL3135            | LADtVEANFT | AtFvPSIRqR | LEndLSGVtL TDtEVtyLMD MCSFDTISTs  |
| <i>A. fumigatus</i> 13073           | LGDEVAANFT | ALFAPdIRAR | aEkhlPGVtL TDEDVVS LMD MCSFDTVarT |
| <i>A. fumigatus</i> 32722           | LGDEVAANFT | ALFAPdIRAR | aEkhlPGVtL TDEDVVS LMD MCSFDTVarT |
| <i>A. fumigatus</i> 58128           | LGDEVAANFT | ALFAPdIRAR | aEkhlPGVtL TDEDVVS LMD MCSFDTVarT |
| <i>A. fumigatus</i> 26906           | LGDEVAANFT | ALFAPdIRAR | aKkhLPGVtL TDEDVVS LMD MCSFDTVarT |
| <i>A. fumigatus</i> 32239           | LGDEVEANFT | ALFAPAIRAR | IEkhLPGVQL TDDDVS LMD MCSFDTVarT  |
| <i>E. nidulans</i>                  | rADEIEANFT | AIMGPPIRkr | LEndLPGIKL TNENViY LMD MCSFDTMarT |
| <i>T. thermophilus</i>              | gGHDaQEKFA | kqFAPAILEK | IKDhLPGVDL AvsDVpyLMD LCPFETLArn  |
| <i>T. lanuginosus</i>               | .DptqpAEf1 | qVFGPRVlkK | ItkhMPGVNL TLEDVplFMD LCPFDTVGsd  |
| <i>M. thermophila</i>               | IGDDaQDtYl | StFAGPITAR | VNAnLPGaNL TDADtVaLMD LCPFETVAsS  |
| Basidio                             | dSDpqxnXWl | AVFAPPITAR | LNAaaPGaNL TDxDaxNLxx LCPFETVS..  |
| Consensus                           | LGDDVEANFT | AVFAPPIRAR | LEA-LPGVNL TDEDVVNLMD MCPFDTVa-T  |
| Fcp10                               | LGDDVEANFT | AVFAPPIRAR | LEAHLPGVNL TDEDVVNLMD MCPFDTVART  |
|                                     | 251        |            | 300                               |
| <i>A. terreus</i> 9a1               | dD..Aht... | .....LSPF  | CDLFta..tE WtQYNYLlSL dKYYGYGGGN  |
| <i>A. terreus</i> cbs               | dD..Aht... | .....LSPF  | CDLFta..aE WtQYNYLlSL dKYYGYGGGN  |
| <i>A. niger</i> var. <i>awamori</i> | Tv..DTK... | .....LSPF  | CDLFTH..dE WiHYDYlQSL kKYYGHGAGN  |
| <i>A. niger</i> NRRL3135            | Tv..DTK... | .....LSPF  | CDLFTH..dE WiNYDYlQSL kKYYGHGAGN  |
| <i>A. fumigatus</i> 13073           | SD..ASQ... | .....LSPF  | CQLFTH..nE WkKYNYLQSL gKYYGYGAGN  |
| <i>A. fumigatus</i> 32722           | SD..ASQ... | .....LSPF  | CQLFTH..nE WkKYNYLQSL gKYYGYGAGN  |
| <i>A. fumigatus</i> 58128           | SD..ASQ... | .....LSPF  | CQLFTH..nE WkKYNYLQSL gKYYGYGAGN  |
| <i>A. fumigatus</i> 26906           | SD..ASQ... | .....LSPF  | CQLFTH..nE WkKYNYLQSL gKYYGYGAGN  |
| <i>A. fumigatus</i> 32239           | AD..ASE... | .....LSPF  | CAIFTH..nE WkKYDYlQSL gKYYGYGAGN  |
| <i>E. nidulans</i>                  | AH..GTE... | .....LSPF  | CAIFTE..kE WlQYDYlQSL sKYYGYGAGS  |
| <i>T. thermophilus</i>              | ht..DT.... | .....LSPF  | CALStQ..eE WqaYDYlQSL gKYYGnGGGN  |
| <i>T. lanuginosus</i>               | PvlfPrQ... | .....LSPF  | CHLFta..dD WmaYDYlYTL dKYYSHGGGS  |
| <i>M. thermophila</i>               | SsdPaTadag | ggngRPLSPF | CrLFSE..sE WraYDYlQSV gKWYGYGPgn  |
| Basidio                             | .....      | ...xexXsXf | CDLFexxpeE FxaFxYxgdL dKFYGTgYgQ  |
| Consensus                           | SD--ATQ--- | -----LSPF  | CDLFTH---E W-QYDYlQSL -KYYGYGAGN  |
| Fcp10                               | SD..ATQ... | .....LSPF  | CDLFTH..DE WIQYDYlQSL gKYYGYGAGN  |

Fig. 4b

|                                     |            |            |                                  |
|-------------------------------------|------------|------------|----------------------------------|
|                                     | 301        |            | 350                              |
| <i>A. terreus</i> 9a1               | PLGPvQGVGW | aNELMARLTR | A.PVHDHTCv NNTLDASPAT FPLNATLYAD |
| <i>A. terreus</i> cbs               | PLGPvQGVGW | aNELIARLTR | S.PVHDHTCv NNTLDANPAT FPLNATLYAD |
| <i>A. niger</i> var. <i>awamori</i> | PLGPTQGVGY | aNELIARLTH | S.PVHDDTSS NHTLDSNPAT FPLNSTLYAD |
| <i>A. niger</i> NRRL3135            | PLGPTQGVGY | aNELIARLTH | S.PVHDDTSS NHTLDSSPAT FPLNSTLYAD |
| <i>A. fumigatus</i> 13073           | PLGPAQGIGF | tNELIARLTR | S.PVQDHTST NsTLvSNPAT FPLNATMYvD |
| <i>A. fumigatus</i> 32722           | PLGPAQGIGF | tNELIARLTR | S.PVQDHTST NsTLvSNPAT FPLNATMYvD |
| <i>A. fumigatus</i> 58128           | PLGPAQGIGF | tNELIARLTR | S.PVQDHTST NsTLvSNPAT FPLNATMYvD |
| <i>A. fumigatus</i> 26906           | PLGPAQGIGF | tNELIARLTR | S.PVQDHTST NsTLvSNPAT FPLNATMYvD |
| <i>A. fumigatus</i> 32239           | PLGPAQGIGF | tNELIARLTN | S.PVQDHTST NsTLvSNPAT FPLNATMYvD |
| <i>E. nidulans</i>                  | PLGPAQGIGF | tNELIARLTQ | S.PVQDNTST NHTLDSNPAT FPLDrkLYAD |
| <i>T. thermophilus</i>              | PLGPAQGVGF | vNELIARMTH | S.PVQDYTTv NHTLDSNPAT FPLNATLYAD |
| <i>T. lanuginosus</i>               | AFGPSRGVGF | vNELIARMTg | NlPVKDHTTv NHTLDdNPET FPLDvLYAD  |
| <i>M. thermophila</i>               | PLGPTQGVGF | vNELIARLA  | GvPVRDgTST NRTLGDPrT FPLGrPLYAD  |
| Basidio                             | PLGPvQGVGY | iNELIARLTx | qa.VRDNTqT NRTLDSsPxT FPLNrTFYAD |

|              |                   |                   |                   |                   |                   |
|--------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Consensus    | PLGPAQGVGF        | -NELIARLTH        | S-PVQDHTST        | NHTLDSNPAT        | FPLNATLYAD        |
| <b>Fcp10</b> | <b>PLGPAQGVGF</b> | <b>VNELIARLTH</b> | <b>S.PVQDHTST</b> | <b>NHTLDSNPAT</b> | <b>FPLNATLYAD</b> |

|                                     |            |            |                                  |
|-------------------------------------|------------|------------|----------------------------------|
|                                     | 351        |            | 400                              |
| <i>A. terreus</i> 9a1               | FSHDSnLVSI | FWALGLYNGT | aPLSqtTSVE. .SvsQTDGYA AAWTVPFAR |
| <i>A. terreus</i> cbs               | FSHDSnLVSI | FWALGLYNGT | kPLSqtTVE. .ditrTDGYA AAWTVPFAR  |
| <i>A. niger</i> var. <i>awamori</i> | FSDHNGIISI | LFALGLYNGT | kPLSTTTVE. .NitQTDGFS SAWTVPFASR |
| <i>A. niger</i> NRRL3135            | FSDHNGIISI | LFALGLYNGT | kPLSTTTVE. .NitQTDGFS SAWTVPFASR |
| <i>A. fumigatus</i> 13073           | FSDHNSMVSI | FFALGLYNGT | ePLSrTSVE. .SaKElDGYS ASWvVPFGAR |
| <i>A. fumigatus</i> 32722           | FSDHNSMVSI | FFALGLYNGT | gPLSrTSVE. .SaKElDGYS ASWvVPFGAR |
| <i>A. fumigatus</i> 58128           | FSDHNSMVSI | FFALGLYNGT | ePLSrTSVE. .SaKElDGYS ASWvVPFGAR |
| <i>A. fumigatus</i> 26906           | FSDHNSMVSI | FFALGLYNGT | ePLSrTSVE. .SaKElDGYS ASWvVPFGAR |
| <i>A. fumigatus</i> 32239           | FSDHNGMIPI | FFAMGLYNGT | ePLSqtSeE. .StKESNGYS ASWAVPFGAR |
| <i>E. nidulans</i>                  | FSDHNSMISI | FFAMGLYNGT | qPLSmdSVE. .SiQEmDGYA ASWTVPFAR  |
| <i>T. thermophilus</i>              | FSDHNTMtSI | FaALGLYNGT | akLSTTeIK. .SiEETDGYS AAWTVPFGR  |
| <i>T. lanuginosus</i>               | FSDHNTMtGI | FsAMGLYNGT | kPLSTSkIQP pTgAAADGYA ASWTVPFAR  |
| <i>M. thermophila</i>               | FSDHNdMMGV | LgALGaYDgV | pPLdkTA..R rdpEELGGYA ASWAVPFAR  |
| Basidio                             | FSDHNqMVAI | FsAMGLFNqS | aPLdPSxpDP nrt.....Wv TsklVPFSAR |

|              |                   |                   |                   |                   |                  |
|--------------|-------------------|-------------------|-------------------|-------------------|------------------|
| Consensus    | FSDHNTMVSI        | FFALGLYNGT        | -PLSTTSVEP        | -S-EETDGYS        | ASWTVPFAR        |
| <b>Fcp10</b> | <b>FSDHNTMVSI</b> | <b>FFALGLYNGT</b> | <b>KPLSTTSVE.</b> | <b>.SIEETDGYS</b> | <b>ASWTVPFAR</b> |

|                                     |            |            |                                   |
|-------------------------------------|------------|------------|-----------------------------------|
|                                     | 401        |            | 450                               |
| <i>A. terreus</i> 9a1               | AYVEMMQC.. | ra.....    | .....EKEPL VRVLVNDVRM PLHGCPtDKL  |
| <i>A. terreus</i> cbs               | AYIEMMQC.. | ra.....    | .....EKQPL VRVLVNDVRM PLHGCAVDNL  |
| <i>A. niger</i> var. <i>awamori</i> | lyVEMMQC.. | Qa.....    | .....EQEPL VRVLVNDRVV PLHGCPIDaL  |
| <i>A. niger</i> NRRL3135            | lyVEMMQC.. | Qa.....    | .....EQEPL VRVLVNDRVV PLHGCPVDaL  |
| <i>A. fumigatus</i> 13073           | AYfEtMQC.. | Ks.....    | .....EKEPL VRaLINDRVV PLHGCDVDKL  |
| <i>A. fumigatus</i> 32722           | AYfEtMQC.. | Ks.....    | .....EKEPL VRaLINDRVV PLHGCDVDKL  |
| <i>A. fumigatus</i> 58128           | AYfEtMQC.. | Ks.....    | .....EKESL VRaLINDRVV PLHGCDVDKL  |
| <i>A. fumigatus</i> 26906           | AYfEtMQC.. | Ks.....    | .....EKEPL VRaLINDRVV PLHGCDVDKL  |
| <i>A. fumigatus</i> 32239           | AYfEtMQC.. | Ks.....    | .....EKEPL VRaLINDRVV PLHGCAVDKL  |
| <i>E. nidulans</i>                  | AYfELMQC.. | E.....     | .....KKEPL VRVLVNDRVV PLHGCAVDKF  |
| <i>T. thermophilus</i>              | AYIEMMQC.. | Dd.....    | .....sDEPV VRVLVNDRVV PLHGCEVDsL  |
| <i>T. lanuginosus</i>               | AYVELLRC.. | Etetsseeee | EG...EDEPF VRVLVNDRVV PLHGCrVDRW  |
| <i>M. thermophila</i>               | iyVEkMRC.. | sggggggggg | EGrqeKDEeM VRVLVNDVRM TLkGCGaDEr  |
| Basidio                             | mvVERLxCxx | xgtxxxxxxx | xxxxxxxxxxx VRVLVNDaVq PLEfCGgDxd |

|              |                   |                |                   |                   |                   |
|--------------|-------------------|----------------|-------------------|-------------------|-------------------|
| Consensus    | AYVEMMQC--        | E-----         | EG---EKEPL        | VRVLVNDRVV        | PLHGCGVDKL        |
| <b>Fcp10</b> | <b>AYVEMMQC..</b> | <b>EA.....</b> | <b>.....EKEPL</b> | <b>VRVLVNDRVV</b> | <b>PLHGCGVDKL</b> |

Fig. 4c

|                                     | 451        |            | 482             |
|-------------------------------------|------------|------------|-----------------|
| <i>A. terreus</i> 9a1               | GRCKrDAFVA | GLSFAQAG.. | GNWADCF~~~ ~~   |
| <i>A. terreus</i> cbs               | GRCKrDDFVE | GLSFARAG.. | GNWAECE~~~ ~~   |
| <i>A. niger</i> var. <i>awamori</i> | GRCtrDsFVr | GLSFARSG.. | GDWAECsA~~~ ~~  |
| <i>A. niger</i> NRRL3135            | GRCtrDsFVr | GLSFARSG.. | GDWAECFA~~~ ~~  |
| <i>A. fumigatus</i> 13073           | GRCKlNDFVK | GLSWARSG.. | GNWGECEFS~~~ ~~ |
| <i>A. fumigatus</i> 32722           | GRCKlNDFVK | GLSWARSG.. | GNWGECEFS~~~ ~~ |
| <i>A. fumigatus</i> 58128           | GRCKlNDFVK | GLSWARSG.. | GNWGECEFS~~~ ~~ |
| <i>A. fumigatus</i> 26906           | GRCKlNDFVK | GLSWARSG.. | GNWGECEFS~~~ ~~ |
| <i>A. fumigatus</i> 32239           | GRCKlKDFVK | GLSWARSG.. | GNSEQSFS~~~ ~~  |
| <i>E. nidulans</i>                  | GRCtlDDWVE | GLNFARSG.. | GNWktCFTl~ ~~   |
| <i>T. thermophilus</i>              | GRCKrDDFVr | GLSFARqG.. | GNWEGCYAas e~   |
| <i>T. lanuginosus</i>               | GRCRrDEWIK | GLTFARqG.. | GHWDrcF~~~ ~~   |
| <i>M. thermophila</i>               | GmCtlErFIE | SMAFARGN.. | GKWDlCFA~~~ ~~  |
| Basidio                             | GxCtlDAFVE | SqxYARedgq | GDFEKCFAtp xx   |
| Consensus                           | GRCK-DDFVE | GLSFARSG-- | GNWEECFA-- --   |
| Fcp10                               | GRCKRDDFVE | GLSFARSG.. | GNWEECFA.. ..   |

Fig. 4d

CP-1  
 Eco RI M G V F V V L L S I A T L F G S T 17  
 TATATGAATTCATGGGCGTGTTCGTGCTACTGTCCATTGCCACCTTGTTTCGGTTCCA  
 1 -----+-----+-----+-----+-----+-----+ 60  
 ATATACTTAAGTACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGT  
  
 S G T A L G P R G N S H S C D T V D G G 37  
 CATCCGGTACCGCCTTGGGTCCTCGTGGTAATTCTCACTCTTGTGACACTGTTGACGGTG  
 61 -----+-----+-----+-----+-----+-----+ 120  
 GTAGGCCATGGCGGAACCCAGGAGCACCATTAAGAGTGAGAACACTGTGACAACTGCCAC  
 CP-2  
 CP-3.10  
 Y Q C F P E I S H L W G Q Y S P F F S L 57  
 GTTACCAATGTTTCCCAGAAATTTCTCACTTGTGGGGTCAATACTCTCCATTCTTCTCTT  
 121 -----+-----+-----+-----+-----+-----+ 180  
 CAATGGTTACAAAGGGTCTTTAAAGAGTGAACACCCAGTTATGAGAGGTAAGAAGAGAA  
  
 A D E S A I S P D V P K G C R V T F V Q 77  
 TGGCTGACGAATCTGCTATTTCTCCAGACGTTCCAAAGGGTGTAGAGTTACTTTCGTTT  
 181 -----+-----+-----+-----+-----+-----+ 240  
 ACCGACTGCTTAGACGATAAAGAGGTCTGCAAGGTTTCCCGACATCTCAATGAAAGCAAG  
 CP-4.10  
 CP-5.10  
 V L S R H G A R Y P T S S K S K K Y S A 97  
 AAGTTTTGTCTAGACACGGTGCTAGATACCCAACTTCTTCTAAGTCTAAGAAGTACTCTG  
 241 -----+-----+-----+-----+-----+-----+ 300  
 TTCAAAACAGATCTGTGCCACGATCTATGGGTTGAAGAAGATTCAGATTCTTCATGAGAC  
  
 L I E A I Q K N A T A F K G K Y A F L K 117  
 CTTTGATTGAAGCTATTCAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGA  
 301 -----+-----+-----+-----+-----+-----+ 360  
 GAAACTAACTTCGATAAGTTTCTTTCGCGATGACGAAAGTTCCCATTTCATGCGAAAGAACT  
 CP-6  
 CP-7.10  
 T Y N Y T L G A D D L T P F G E Q Q M V 137  
 AGACTTACAAC TACACTTTGGGTGCTGACGACTTGACTCCATTGCGGTGAACAACAAATGG  
 361 -----+-----+-----+-----+-----+-----+ 420  
 TCTGAATGTTGATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTGTTGTTTACC  
  
 N S G I K F Y R R Y K A L A R K I V P F 157  
 TTAAC TCTGGTATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCAT  
 421 -----+-----+-----+-----+-----+-----+ 480  
 AATTGAGACCATAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTA  
 CP-8.10  
 CP-9.10  
 V R A S G S D R V I A S A E K F I E G F 177  
 TCGTTAGAGCTTCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTTCATTGAAGGTT  
 481 -----+-----+-----+-----+-----+-----+ 540  
 AGCAATCTCGAAGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACTTCAA  
  
 Q S A K L A D P G A N P H Q A S P V I N 197  
 TCCAATCTGCTAAGTTGGCTGACCCAGGTGCTAACCCACACCAAGCTTCTCCAGTTATTA  
 541 -----+-----+-----+-----+-----+-----+ 600  
 AGGTTAGACGATTCAACCGACTGGGTCCACGATTGGGTGTGTTTCAAGAGGTCAATAAT

Fig. 5a

15/56

CP-10.10

CP-11.10

V I I P E G A G Y N N T L D H G L C T A 217  
**ACGTTATTATTCCAGAAGGTGCTGGTTACAACAACACTTTGGACCACGGTTTGTGTACTG**  
 601 -----+-----+-----+-----+-----+-----+ 660  
**TGCAATAATAAGGTCCTCCACGACCAATGTTGTTGTGAAACCTGGTGCCAAACACATGAC**

F E E S E L G D D V E A N F T A V F A P 237  
**CFTTCGAAGAATCTGAATTGGGTGACGACGTTGAAGCTAACTTCACTGCTGTTTTCGCTC**  
 661 -----+-----+-----+-----+-----+-----+ 720  
**GAAAGCTTCTTAGACTTAACCCACTGCTGCAACTTCGATTGAAGTGACGACAAAAGCGAG**

CP-12.10

P I R A R L E A H L P G V N L T D E D V 257  
**CACCTATTAGAGCTAGATTGGAAGCTCACTTGCCAGGTGTTAACTTGACTGACGAAGACG**  
 721 -----+-----+-----+-----+-----+-----+ 780  
**GTGGATAATCTCGATCTAACCTTCGAGTGAACGGTCCACAATTGAACTGACTGCTTCTGC**

CP-13.10

V N L M D M C P F D T V A R T S D A T Q 277  
**TTGTTAACTTGATGGACATGTGTCCATTGCACACTGTTGCTAGAACTTCTGACGCTACTC**  
 781 -----+-----+-----+-----+-----+-----+ 840  
**AACAATTGAACTACCTGTACACAGGTAAGCTGTGACAACGATCTTGAAGACTGCGATGAG**

L S P F C D L F T H D E W I Q Y D Y L Q 297  
**AATTGTCTCCATTCTGTGACTTGTTCACTCACGACGAATGGATTCAATACGACTACTTGC**  
 841 -----+-----+-----+-----+-----+-----+ 900  
**TTAACAGAGGTAAGACACTGAACAAGTGAGTGCTGCTTACCTAAGTTATGCTGATGAACG**

CP-14.10

CP-15.10

S L G K Y Y G Y G A G N P L G P A Q G V 317  
**AATCTTTGGGTAAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTG**  
 901 -----+-----+-----+-----+-----+-----+ 960  
**TTAGAAACCCATTTCATGATGCCAATGCCACGACCATTGGGTAACCCAGGTGAGTTCCAC**

G F V N E L I A R L T H S P V Q D H T S 337  
**TTGGTTTCGTTAACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTT**  
 961 -----+-----+-----+-----+-----+-----+ 1020  
**AACCAAAGCAATTGCTTAACTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGTGAA**

CP-16.10

CP-17.10

T N H T L D S N P A T F P L N A T L Y A 357  
**CTACTAACCACACTTTGGACTCTAACCCAGCTACTTTCCCATTTGAACGCTACTTTGTACG**  
 1021 -----+-----+-----+-----+-----+-----+ 1080  
**GATGATTGGTGTGAAACCTGAGATTGGGTGATGAAAGGGTAACCTTGCGATGAAACATGC**

D F S H D N T M V S I F F A L G L Y N G 377  
**CTGACTTCTCTCAGACAACACTATGGTTTCTATTTTCTTCGCTTTGGGTTTGTACAACG**  
 1081 -----+-----+-----+-----+-----+-----+ 1140  
**GACTGAAGAGAGTGCTGTTGTGATACCAAAGATAAAAGAAGCGAAACCCAAACATGTTGC**

CP-18.10

CP-19.10

T K P L S T T S V E S I E E T D G Y A A 397  
**GTACTAAGCCATTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACGCTG**  
 1141 -----+-----+-----+-----+-----+-----+ 1200  
**CATGATTGCGTAACAGATGATGAAGACAACCTTAGATAACTTCTTTGACTGCCAATGCGAC**

Fig. 5b

00210" 5328460

16/56

S W T V P F A A R A Y V E M M Q C E A E 417  
CTTCTTGGA**CTGTTCCATT**CGCTGCTAGAGCTTACGTTGAAATGATGCAATGTGAAGCTG  
1201 -----+-----+-----+-----+-----+-----+ 1260  
GAAGAACC**TGACAAGGTAAGCCACGATCTCGAATGCAACTTTACTACGTTACACTTCGAC**  
CP-20.10  
CP-21.10  
K E P L V R V L V N D R V V P L H G C G 437  
AAAAGGAACCATTGGTTAGAGTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTG  
1261 -----+-----+-----+-----+-----+-----+ 1320  
TTTTCTTGGA**TAACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACAC**  
V D K L G R C K R D D F V E G L S F A R 457  
GTGTTGACAAGTTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTA  
1321 -----+-----+-----+-----+-----+-----+ 1380  
CACAACTGTTCAACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGAT  
CP-22.10  
S G G N W E E C F A \* Eco RI 467  
GATCTGGTGGTA**ACTGGGAAGAAATGTTTCGCTTAAGAATT**CATATA  
1381 -----+-----+-----+-----+-----+ 1426  
CTAGACCACCATTGACCCTTCTTACAAAGCGAATTCTTAAGTATAT

Fig. 5c

bioRxiv preprint doi: <https://doi.org/10.1101/092150>; this version posted September 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 50

*P. involutus* (phyA1) ~~~~~ ~FPipeseqR nWSPYSPYFP LAEyKA... pPaGCQInqV  
*P. involutus* (phyA2) ~~~~~ ~FsipeseqR nWSPYSPYFP LAEyKA... pPaGCeInqV  
*T. pubescens* ~~~~~ ~LDvtRDVqQ sWSmYSPYFP aAtyvA... pPaSCQInqV  
*A. pediades* ~~~~~ ~pfppPQIQD sWAaYTPYYP VqAyTP... pPKDCKITqV  
*P. lycii* ~~~~~ ~LPipAQnTs nWGPYdPFFP VEpyAA... pPEGCTVTqV  
*A. terreus* 9a1 KhSDCNSVDh GYQCfPELSh kWGLYAPYFS LqDESFPFLD VPEDCHITFV  
*A. terreus* cbs NhSDCtSVDr GYQCfPELSh kWGLYAPYFS LqDESFPFLD VPDDCHITFV  
*A. niger* var. *awamori* NqSTCDTVDq GYQCfSEtSH LWGQYAPFFS LANESAISPD VPAGCRVTfFa  
*A. niger* T213 NqSSCDTVDq GYQCfSEtSH LWGQYAPFFS LANESvISPD VPAGCRVTfFa  
*A. niger* NRRL3135 NqSSCDTVDq GYQCfSEtSH LWGQYAPFFS LANESvISPE VPAGCRVTfFa  
*A. fumigatus* ATCC13073 GSkSCDTVD1 GYQCSPAtSH LWGQYSPFFS LEDELsvSSK LPKDCRITLV  
*A. fumigatus* ATCC32722 GSkSCDTVD1 GYQCSPAtSH LWGQYSPFFS LEDELsvSSK LPKDCRITLV  
*A. fumigatus* ATCC58128 GSkSCDTVD1 GYQCSPAtSH LWGQYSPFFS LEDELsvSSK LPKDCRITLV  
*A. fumigatus* ATCC26906 GSkSCDTVD1 GYQCSPAtSH LWGQYSPFFS LEDELsvSSK LPKDCRITLV  
*A. fumigatus* ATCC32239 GSkACDTVE1 GYQCSPGtSH LWGQYSPFFS LEDELsvSSD LPKDCRVTFV  
*E. nidulans* QNHSCNTaDg GYQCfPNVSH VWGQYSPYFS IEQESAISed VPhGCeVTFV  
*T. thermophilus* DSHSCNTVEg GYQCfPEISH sWGQYSPFFS LADQSEISPD VPQNCKITFV  
*T. lanuginosus* ~~~~~ ~~~~~nvDIAR hWGQYSPFFS LAEvSEISPA VPKGCRVeFfV  
*M. thermophila* ESRPCDTpD1 GFQCgTAISH FWGQYSPYFS VPSElDaS.. IPDDCeVTFa

**Consensus Seq. 11** NSHSCDTVD- GYQC-PEISH LWGQYSPFFS LADESAISPD VPKGCRVTfV

51 100

*P. involutus* (phyA1) NIIqRHGARF PTSGaTtRik AgLtKLQgvq nftDAKFNF1 KSFKYdLGns  
*P. involutus* (phyA2) NIIqRHGARF PTSGaAtRik AgLsKLQsvq nftDPKFDFI KSFTYdLGTs  
*T. pubescens* HIIqRHGARF PTSGaAKRiq TaVAKLKaaS nytDPLLAfV tnYtYSLGqD  
*A. pediades* NIIqRHGARF PTSGaGtRiq AaVKKLQsak TytDPRLDfL tnYtYTLGhD  
*P. lycii* NLIqRHGARW PTSGarsRqv AaVAKIQmar PftDPKYEFL NdfVYkFGvA  
*A. terreus* 9a1 QVLaRHGARS PThSKTKaYA AtIAaIQKSA TaFpGKYAFL QSYNYSLDSE  
*A. terreus* cbs QVLaRHGARS PtdSKTKaYA AtIAaIQKNA TaLpGKYAFL KSYNYSMGSE  
*A. niger* var. *awamori* QVLSRHGARY PTeSKGKKYS ALIEeIQQNv TtFDGKYAFL KTYNYSLGAD  
*A. niger* T213 QVLSRHGARY PTeSKGKKYS ALIEeIQQNv TtFDGKYAFL KTYNYSLGAD  
*A. niger* NRRL3135 QVLSRHGARY PtdSKGKKYS ALIEeIQQNA TtFDGKYAFL KTYNYSLGAD  
*A. fumigatus* ATCC13073 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL KTYNYTLGAD  
*A. fumigatus* ATCC32722 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL KTYNYTLGAD  
*A. fumigatus* ATCC58128 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL KTYNYTLGAD  
*A. fumigatus* ATCC26906 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL KTYNYTLGAD  
*A. fumigatus* ATCC32239 QVLSRHGARY PTASKSKKYk kLVtaIQKNA TeFKGKFAFL ETYNYTLGAD  
*E. nidulans* QVLSRHGARY PTeSKSKaYS GLIEaIQKNA TsFwGQYAFL ESYNYTLGAD  
*T. thermophilus* QLLSRHGARY PTSSKTELYS qLISRIQKtA TaYKGyYAFL KdYrYqLGAN  
*T. lanuginosus* QVLSRHGARY PTAhKSEvYA ELLQRIQDtA TeFKGDFAFL RdYaYhLGAD  
*M. thermophila* QVLSRHGARA PtlkRaasYv DLIDRIHhGA isYgPgYEFL RTYDYTLGAD

**Consensus Seq. 11** QVLSRHGARY PTSSKSKKYS ALIERIQKNA T-FKGKYAFL KTYNYTLGAD

Fig. 6a

|                                     |                                                               |  |     |
|-------------------------------------|---------------------------------------------------------------|--|-----|
|                                     | 101                                                           |  | 150 |
| <i>P. involutus</i> (phyA1)         | DLvPFGAaQs fDAGqEaFaR YskLvSKNnL PFIRAdGSDR VVDSAtNWtA        |  |     |
| <i>P. involutus</i> (phyA2)         | DLvPFGAaQs fDAGLEvFaR YskLvSsDnL PFIRSdGSDR VVDtAtNWtA        |  |     |
| <i>T. pubescens</i>                 | sLveLGAtQs sEAGqEaFtR YsSLvSaDeL PFVRASGSDR VVATANNWtA        |  |     |
| <i>A. pediades</i>                  | DLvPFGAlQs sQAGeEtFQR YsfLvSKEnL PFVRASSNR VVDSAtNWtE         |  |     |
| <i>P. lycii</i>                     | DLlPFGANQs hQTGTDMYtR YsTLfEgGdV PFVRAAGdQR VVDSStNWtA        |  |     |
| <i>A. terreus</i> 9a1               | ELTPFGrNQL rDlGaQFYeR YNAL.TRHIn PFVRATDAsR VhESAeKFVE        |  |     |
| <i>A. terreus</i> cbs               | NLTPFGrNQL qDlGaQFYRR YDTL.TRHIn PFVRAADSsR VhESAeKFVE        |  |     |
| <i>A. niger</i> var. <i>awamori</i> | DLTPFGEQEL VNSGIKFYQR YESL.TRNII PFIRSSGSsR VIASGEKFIE        |  |     |
| <i>A. niger</i> T213                | DLTPFGEQEL VNSGIKFYQR YESL.TRNII PFIRSSGSsR VIASGEKFIE        |  |     |
| <i>A. niger</i> NRRL3135            | DLTPFGEQEL VNSGIKFYQR YESL.TRNIV PFIRSSGSsR VIASGKKFIE        |  |     |
| <i>A. fumigatus</i> ATCC13073       | DLTPFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE        |  |     |
| <i>A. fumigatus</i> ATCC32722       | DLTPFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE        |  |     |
| <i>A. fumigatus</i> ATCC58128       | DLTPFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE        |  |     |
| <i>A. fumigatus</i> ATCC26906       | DLTAFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE        |  |     |
| <i>A. fumigatus</i> ATCC32239       | DLTPFGEQQM VNSGIKFYQK YKAL.AgSVV PFIRSSGSDR VIASGEKFIE        |  |     |
| <i>E. nidulans</i>                  | DLTiFGENQM VDSGaKFYRR YKnL.ARKnt PFIRASGSDR VVASAEKFIn        |  |     |
| <i>T. thermophilus</i>              | DLTPFGENQM IQlGIKFYnH YKSL.ARNvV PFVRCsGSDR VIASGrIFIE        |  |     |
| <i>T. lanuginosus</i>               | NLTRFGEEQM MESGrQFYHR YREq.AREIV PFVRAAGSAR VIASAEffNr        |  |     |
| <i>M. thermophila</i>               | ELTRtGQQQM VNSGIKFYRR YRAL.ARKS I PFVRTAGqDR VVhSAENftQ       |  |     |
| <b>Consensus Seq. 11</b>            | <b>DLTPFGENQM VNSGIKFYRR YKAL-ARNIV PFVRASGSDR VIASAEKFIE</b> |  |     |
|                                     | 151                                                           |  | 200 |
| <i>P. involutus</i> (phyA1)         | GFaSA..... ..shNtvqPk LNLILPQ..T gNDTLEDNMC PAaGD.....        |  |     |
| <i>P. involutus</i> (phyA2)         | GFaSA..... ..srNaiqPk LDLILPQ..T gNDTLEDNMC PAaGE.....        |  |     |
| <i>T. pubescens</i>                 | GFaLA..... ..ssNsITPV LSVIIE..A gNDTLDDNMC PAaGD.....         |  |     |
| <i>A. pediades</i>                  | GFsAA..... ..shHvLNPI LfVILSE..S LNDTLDDAMC PnaGs.....        |  |     |
| <i>P. lycii</i>                     | GFgdA..... ..sgEtvLpt LQVVLQE..E gNcTLcNNMC PnevD.....        |  |     |
| <i>A. terreus</i> 9a1               | GFQTARqDDh hAnpHQPSPr VDVAIPEGSA YNNTLEHSLC TAFES...ST        |  |     |
| <i>A. terreus</i> cbs               | GFQNARqGDP hAnpHQPSPr VDVVIPEGTA YNNTLEHSIC TAFEa...ST        |  |     |
| <i>A. niger</i> var. <i>awamori</i> | GFQSTKLkDP rAqpgQSSPk IDVVISeASS sNNTLDpGtC TvFED...Se        |  |     |
| <i>A. niger</i> T213                | GFQSTKLkDP rAqpgQSSPk IDVVISeASS sNNTLDpGtC TvFED...Se        |  |     |
| <i>A. niger</i> NRRL3135            | GFQSTKLkDP rAqpgQSSPk IDVVISeASS sNNTLDpGtC TvFED...Se        |  |     |
| <i>A. fumigatus</i> ATCC13073       | GFQqAKLADP gAt.NRAAPa ISVIIPeSeT FNNTLDHGVC TkFEa...Sq        |  |     |
| <i>A. fumigatus</i> ATCC32722       | GFQqAKLADP gAt.NRAAPa ISVIIPeSeT FNNTLDHGVC TkFEa...Sq        |  |     |
| <i>A. fumigatus</i> ATCC58128       | GFQqAKLADP gAt.NRAAPa ISVIIPeSeT FNNTLDHGVC TkFEa...Sq        |  |     |
| <i>A. fumigatus</i> ATCC26906       | GFQqAKLADP gAt.NRAAPa ISVIIPeSeT FNNTLDHGVC TkFEa...Sq        |  |     |
| <i>A. fumigatus</i> ATCC32239       | GFQqANVADP gAt.NRAAPV ISVIIPeSeT YNNTLDHSVC TnFEa...Se        |  |     |
| <i>E. nidulans</i>                  | GFRkAQlhdh g.s.gQATPV VNVIPEidG FNNTLDHStC vSFEN...de         |  |     |
| <i>T. thermophilus</i>              | GFQSAKVlDP hSdKHDApPt INVIIeGPS YNNTLDtGsC PvFED...SS         |  |     |
| <i>T. lanuginosus</i>               | GFQdAKdrDP rSnkDQaEPV INVIISeETG sNNTLDgltC PAaEE...AP        |  |     |
| <i>M. thermophila</i>               | GFHSALLADR gStvRPTlPy dmVVIPEtAG aNNTLHNDLC TAFEegpyST        |  |     |
| <b>Consensus Seq. 11</b>            | <b>GFQSAKLADP -A--HQASPV INVIIPEGSG YNNTLDHGVC TAFED---ST</b> |  |     |

Fig. 6b

|                                     |                                                               |  |     |
|-------------------------------------|---------------------------------------------------------------|--|-----|
|                                     | 201                                                           |  | 250 |
| <i>P. involutus</i> (phyA1)         | .SDpqvnaWl AVafPSItAR LNAAApsvNL TdtdafNLVs LCaFlTVSK.        |  |     |
| <i>P. involutus</i> (phyA2)         | .SDpqvDaWl AsafPSVtAQ LNAAApgANL TDADafNLVs LCPFMtVSK.        |  |     |
| <i>T. pubescens</i>                 | .SDpqvnQWl AqFAPPMtAR LNagaPGaNL TdtdtyNLlt LCPFETVAt.        |  |     |
| <i>A. pediades</i>                  | .SDpqtGiWT SIYGTPIanR LNqqaPGaNI TAADVsnLIp LCAFETivK.        |  |     |
| <i>P. lycii</i>                     | .GDESt.tWl GVfAPnItAR LNAAApsaNL SDsDaLtLMD MCPFDtLSs.        |  |     |
| <i>A. terreus</i> 9a1               | VGDDAvANFT AVFAPAIaQR LEAdLPGVQL StDDVVNLMA MCPFETVSlT        |  |     |
| <i>A. terreus</i> cbs               | VGDAADNFT AVFAPAIaQR LEAdLPGVQL SADDVVNLMA MCPFETVSlT         |  |     |
| <i>A. niger</i> var. <i>awamori</i> | LADtveANFT AtFAPSIRqR LEndLSGVtL TdteVtyLMD MCSFDtISs         |  |     |
| <i>A. niger</i> T213                | LADtveANFT AtFAPSIRqR LEndLSGVtL TdteVtyLMD MCSFDtISs         |  |     |
| <i>A. niger</i> NRRL3135            | LADtveANFT AtFvPSIRqR LEndLSGVtL TdteVtyLMD MCSFDtISs         |  |     |
| <i>A. fumigatus</i> ATCC13073       | LGDEvAANFT ALFAPdIRAR aEkhlPGVtL TDEDVvSLMD MCSFDtVART        |  |     |
| <i>A. fumigatus</i> ATCC32722       | LGDEvAANFT ALFAPdIRAR aEkhlPGVtL TDEDVvSLMD MCSFDtVART        |  |     |
| <i>A. fumigatus</i> ATCC58128       | LGDEvAANFT ALFAPdIRAR aEkhlPGVtL TDEDVvSLMD MCSFDtVART        |  |     |
| <i>A. fumigatus</i> ATCC26906       | LGDEvAANFT ALFAPdIRAR aKkhLPGVtL TDEDVvSLMD MCSFDtVART        |  |     |
| <i>A. fumigatus</i> ATCC32239       | LGDEvAANFT ALFAPAIRAR IEkhLPGVQL TDDDVvSLMD MCSFDtVART        |  |     |
| <i>E. nidulans</i>                  | rADEiEANFT AIMGPPIRkR LEndLPGIKL TNENViyLMD MCSFDtMART        |  |     |
| <i>T. thermophilus</i>              | gGHDAQEKFA kqFAPAILEK IKDhLPGVDL AvsDVpyLMD LCPFETLARN        |  |     |
| <i>T. lanuginosus</i>               | .DptqpAEFl qVFGPRVlkK ItkhMPGVNL TLEDVplFMD LCPFDtVGsd        |  |     |
| <i>M. thermophila</i>               | IGDDAQDtYl StFAGPItAR VNANLPGaNL TDADtVaLMD LCPFETVAsS        |  |     |
| <b>Consensus Seq. 11</b>            | <b>LGDDAEANFT AVFAPPiRAR LEA-LPGVNL TDEDVVNLMD MCPFDtVART</b> |  |     |
|                                     | 251                                                           |  | 300 |
| <i>P. involutus</i> (phyA1)         | ..... .ekkSdF CtLFegIPGs FeaFAYggdL dKFYGTGyGQ                |  |     |
| <i>P. involutus</i> (phyA2)         | ..... .eqkSdF CtLFegIPGs FeaFAYagdL dKFYGTGyGQ                |  |     |
| <i>T. pubescens</i>                 | ..... .errSeF CDiYeelqAE .daFAYnadL dKFYGTGyGQ                |  |     |
| <i>A. pediades</i>                  | ..... .etpSPF CNLF..TPEE FaQFEYFgdL dKFYGTGyGQ                |  |     |
| <i>P. lycii</i>                     | ..... .gnaSPF CDLF..TAAE YvsYEYYdL dKYYGTGPGN                 |  |     |
| <i>A. terreus</i> 9a1               | dD..Aht... ..LSPF CDLF..TatE WtQYNYLlSL dKYYGYGGGN            |  |     |
| <i>A. terreus</i> cbs               | dD..Aht... ..LSPF CDLF..TAAE WtQYNYLlSL dKYYGYGGGN            |  |     |
| <i>A. niger</i> var. <i>awamori</i> | Tv..DTK... ..LSPF CDLF..ThDE WiHYDYlQSL kKYYGHGAGN            |  |     |
| <i>A. niger</i> T213                | Tv..DTK... ..LSPF CDLF..ThDE WiHYDYlRSL kKYYGHGAGN            |  |     |
| <i>A. niger</i> NRRL3135            | Tv..DTK... ..LSPF CDLF..ThDE WiNYDYlQSL kKYYGHGAGN            |  |     |
| <i>A. fumigatus</i> ATCC13073       | SD..ASQ... ..LSPF CQLF..ThNE WkKYNyLQSL gKYYGYGAGN            |  |     |
| <i>A. fumigatus</i> ATCC32722       | SD..ASQ... ..LSPF CQLF..ThNE WkKYNyLQSL gKYYGYGAGN            |  |     |
| <i>A. fumigatus</i> ATCC58128       | SD..ASQ... ..LSPF CQLF..ThNE WkKYNyLQSL gKYYGYGAGN            |  |     |
| <i>A. fumigatus</i> ATCC26906       | SD..ASQ... ..LSPF CQLF..ThNE WkKYNyLQSL gKYYGYGAGN            |  |     |
| <i>A. fumigatus</i> ATCC32239       | AD..ASE... ..LSPF CAIF..ThNE WkKYDYlQSL gKYYGYGAGN            |  |     |
| <i>E. nidulans</i>                  | AH..GTE... ..LSPF CAIF..TEKE WlQYDYlQSL sKYYGYGAGS            |  |     |
| <i>T. thermophilus</i>              | ht..DT.... ..LSPF CALs..TqEE WqaYDYyQSL gKYYGnGGGN            |  |     |
| <i>T. lanuginosus</i>               | PvlfPrQ... ..LSPF CHLF..TADD WmaYDYyTL dKYYSHGGGS             |  |     |
| <i>M. thermophila</i>               | SsdpATadag ggngRpLSPF CrLF..SEsE WraYDYlQSV gKWYGYGPGN        |  |     |
| <b>Consensus Seq. 11</b>            | <b>SD--ATQ--- -----LSPF CDLF--TADE W-QYDYlQSL -KYYGYGAGN</b>  |  |     |

Fig. 6c

|                                     |                               |                   |                              |
|-------------------------------------|-------------------------------|-------------------|------------------------------|
|                                     | 301                           |                   | 350                          |
| <i>P. involutus</i> (phyA1)         | eLGPvQGVGY vNELIARLTN         | S.AVRDNTqT        | NRTLDAASPvT FPLNkTFYAD       |
| <i>P. involutus</i> (phyA2)         | ALGPvQGVGY iNELIARLTN         | S.AVNDNTqT        | NRTLDAaPDT FPLNkTMYAD        |
| <i>T. pubescens</i>                 | PLGPvQGVGY iNELIARLTa         | q.nVsDHTqT        | NsTLdSSPET FPLNrTLYAD        |
| <i>A. pediades</i>                  | PLGPvQGVGY iNELIARLTE         | m.PVRDNTqT        | NRTLdSSPlT FPLDrSIYAD        |
| <i>P. lycii</i>                     | ALGPvQGVGY vNELIARLTg         | q.AVRDETqT        | NRTLdSDPAT FPLNrTFYAD        |
| <i>A. terreus</i> 9a1               | PLGPvQGVGW aNELMARLTR         | A.PVHDHTCv        | NNTLDASPAT FPLNATLYAD        |
| <i>A. terreus</i> cbs               | PLGPvQGVGW aNELIARLTR         | S.PVHDHTCv        | NNTLDANPAT FPLNATLYAD        |
| <i>A. niger</i> var. <i>awamori</i> | PLGPTQGVGY aNELIARLTH         | S.PVHDDTSS        | NHTLdSNPAT FPLNSTLYAD        |
| <i>A. niger</i> T213                | PLGPTQGVGY aNELIARLTH         | S.PVHDDTSS        | NHTLdSNPAT FPLNSTLYAD        |
| <i>A. niger</i> NRRL3135            | PLGPTQGVGY aNELIARLTH         | S.PVHDDTSS        | NHTLdSSPAT FPLNSTLYAD        |
| <i>A. fumigatus</i> ATCC13073       | PLGPAQGIGF tNELIARLTR         | S.PVQDHTST        | NsTLvSNPAT FPLNATMYvD        |
| <i>A. fumigatus</i> ATCC32722       | PLGPAQGIGF tNELIARLTR         | S.PVQDHTST        | NsTLvSNPAT FPLNATMYvD        |
| <i>A. fumigatus</i> ATCC58128       | PLGPAQGIGF tNELIARLTR         | S.PVQDHTST        | NsTLvSNPAT FPLNATMYvD        |
| <i>A. fumigatus</i> ATCC26906       | PLGPAQGIGF tNELIARLTR         | S.PVQDHTST        | NsTLvSNPAT FPLNATMYvD        |
| <i>A. fumigatus</i> ATCC32239       | PLGPAQGIGF tNELIARLTN         | S.PVQDHTST        | NsTLdSDPAT FPLNATIYvD        |
| <i>E. nidulans</i>                  | PLGPAQGIGF tNELIARLTQ         | S.PVQDNTST        | NHTLdSNPAT FPLDrkLYAD        |
| <i>T. thermophilus</i>              | PLGPAQGVGF vNELIARMTg         | S.PVQDYTTv        | NHTLdSNPAT FPLNATLYAD        |
| <i>T. lanuginosus</i>               | AFGSPSRGVGF vNELIARMTg        | NlPVKDHTTv        | NHTLdNPET FPLDAvLYAD         |
| <i>M. thermophila</i>               | PLGPTQGVGF vNELIARLA.         | GvPVRDgTST        | NRTLdGDPrT FPLGrPLYAD        |
| <b>Consensus Seq. 11</b>            | <b>PLGPAQGVGF -NELIARLTH</b>  | <b>S-PVQDHTST</b> | <b>NHTLdSNPAT FPLNATLYAD</b> |
|                                     | 351                           |                   | 400                          |
| <i>P. involutus</i> (phyA1)         | FSHDNlMVAV FsAMGLFrqP         | aPLSTsvpNP        | wrt.....Wr TSSlVPFSGR        |
| <i>P. involutus</i> (phyA2)         | FSHDNlMVAV FsAMGLFrqS         | aPLSTSTpDP        | nrt.....Wl TSSvVPFSAR        |
| <i>T. pubescens</i>                 | FSHDNqMVAI FsAMGLFNqS         | aPLdPTTpDP        | art.....Fl vkkiVPFSAR        |
| <i>A. pediades</i>                  | LShDNqMIAI FsAMGLFNqS         | sPLdPSfpNP        | krt.....Wv TSRLtPFSAR        |
| <i>P. lycii</i>                     | FSHDNTMVPI FaALGLFNAT         | a.LdPlkpDe        | nr1.....Wv DSk1VPFSGH        |
| <i>A. terreus</i> 9a1               | FSHDSnLVSI FWALGLYNGT         | aPLSqTSVES        | Vs..QTDGYA AAWTVPFAAR        |
| <i>A. terreus</i> cbs               | FSHDSnLVSI FWALGLYNGT         | KPLSqTTVEd        | It..rTDGYA AAWTVPFAAR        |
| <i>A. niger</i> var. <i>awamori</i> | FSHDNGIISI LFALGLYNGT         | KPLSTTTVEN        | It..QTDGFS SAWTVPFASR        |
| <i>A. niger</i> T213                | FSHDNGIISI LFALGLYNGT         | KPLSTTTVEN        | It..QTDGFS SAWTVPFASR        |
| <i>A. niger</i> NRRL3135            | FSHDNGIISI LFALGLYNGT         | KPLSTTTVEN        | It..QTDGFS SAWTVPFASR        |
| <i>A. fumigatus</i> ATCC13073       | FSHDNSMVISI FFALGLYNGT        | EPLSrTSVES        | ak..ElDGYS ASWvVPFGAR        |
| <i>A. fumigatus</i> ATCC32722       | FSHDNSMVISI FFALGLYNGT        | gPLSrTSVES        | ak..ElDGYS ASWvVPFGAR        |
| <i>A. fumigatus</i> ATCC58128       | FSHDNSMVISI FFALGLYNGT        | EPLSrTSVES        | ak..ElDGYS ASWvVPFGAR        |
| <i>A. fumigatus</i> ATCC26906       | FSHDNSMVISI FFALGLYNGT        | EPLSrTSVES        | ak..ElDGYS ASWvVPFGAR        |
| <i>A. fumigatus</i> ATCC32239       | FSHDNGMIPI FFAMGLYNGT         | EPLSqTSeES        | tk..ESNGYS ASWAVPFGAR        |
| <i>E. nidulans</i>                  | FSHDNSMISI FFAMGLYNGT         | QPLSmdSVES        | Iq..EmDGYA ASWTVPFGAR        |
| <i>T. thermophilus</i>              | FSHDNTMtSI FaALGLYNGT         | akLSTTeIKS        | Ie..ETDGYS AAWTVPFGR         |
| <i>T. lanuginosus</i>               | FSHDNTMtGI FsAMGLYNGT         | KPLSTSkIQP        | ptgaAADGYA ASWTVPFAAR        |
| <i>M. thermophila</i>               | FSHDNdMMGV LgALGaYDGv         | pPLdkTArrd        | ..peElGGYA ASWAVPFAAR        |
| <b>Consensus Seq. 11</b>            | <b>FSHDNTMVISI FFALGLYNGT</b> | <b>KPLSTTSVES</b> | <b>I---ETDGYS ASWTVPFAAR</b> |

Fig. 6d

|                                     |                       |                              |                              |
|-------------------------------------|-----------------------|------------------------------|------------------------------|
|                                     | 401                   |                              | 450                          |
| <i>P. involutus</i> (phyA1)         | mvVERLsC.. fGt.....   | .....Tk VRVLVQDQVq           | PLEfCGgDRn                   |
| <i>P. involutus</i> (phyA2)         | maVERLsC.. AGt.....   | .....Tk VRVLVQDQVq           | PLEfCGgDQd                   |
| <i>T. pubescens</i>                 | mvVERLDC.. GGa.....   | .....Qs VRLLVNDaVq           | PLafCGaDts                   |
| <i>A. pediades</i>                  | mvTErLLCQr DGtGsGGpsr | imrNgnvQTF VRILVNDaLq        | PLkfCGgDmd                   |
| <i>P. lycii</i>                     | mtVEkLaC.. .....      | .....sgKea VRVLVNDaVq        | PLEfCGg.vd                   |
| <i>A. terreus</i> 9a1               | AYVEMMQCrA .....      | ..EK...EPL VRVLVNDRVM        | PLHGCPtDKL                   |
| <i>A. terreus</i> cbs               | AYIEMMQCrA .....      | ..EK...QPL VRVLVNDRVM        | PLHGCAVDNL                   |
| <i>A. niger</i> var. <i>awamori</i> | lYVEMMQCQA .....      | ..EQ...EPL VRVLVNDRVV        | PLHGCPIDaL                   |
| <i>A. niger</i> T213                | lYVEMMQCQA .....      | ..EQ...EPL VRVLVNDRVV        | PLHGCPIDaL                   |
| <i>A. niger</i> NRRL3135            | lYVEMMQCQA .....      | ..EQ...EPL VRVLVNDRVV        | PLHGCPVDaL                   |
| <i>A. fumigatus</i> ATCC13073       | AYfetMQCKS .....      | ..EK...EPL VRaLINDRVV        | PLHGCDVDKL                   |
| <i>A. fumigatus</i> ATCC32722       | AYfetMQCKS .....      | ..EK...EPL VRaLINDRVV        | PLHGCDVDKL                   |
| <i>A. fumigatus</i> ATCC58128       | AYfetMQCKS .....      | ..EK...ESL VRaLINDRVV        | PLHGCDVDKL                   |
| <i>A. fumigatus</i> ATCC26906       | AYfetMQCKS .....      | ..EK...EPL VRaLINDRVV        | PLHGCDVDKL                   |
| <i>A. fumigatus</i> ATCC32239       | AYfetMQCKS .....      | ..EK...EPL VRaLINDRVV        | PLHGCAVDKL                   |
| <i>E. nidulans</i>                  | AYfELMQCE. ....       | ..KK...EPL VRVLVNDRVV        | PLHGCAVDKF                   |
| <i>T. thermophilus</i>              | AYIEMMQCDD .....      | ..sD...EPV VRVLVNDRVV        | PLHGCEVDsL                   |
| <i>T. lanuginosus</i>               | AYVELLRCEt ETsSeEEeEG | ..ED...EPF VRVLVNDRVV        | PLHGCrVDRw                   |
| <i>M. thermophila</i>               | iYVEkMRCSg GGgGgGGgEG | ..rQekdEeM VRVLVNDRVM        | TLkGCGaDEr                   |
| <b>Consensus Seq. 11</b>            | <b>AYVEMMQCEA</b>     | <b>GG-G-GG-EG --EK---EPL</b> | <b>VRVLVNDRVV PLHGCGVDKL</b> |
|                                     | 451                   |                              | 482                          |
| <i>P. involutus</i> (phyA1)         | GlCtLAKFVE SqTFARSDga | GDFEKCFAts a~                |                              |
| <i>P. involutus</i> (phyA2)         | GlCaLDKFVE SqAYARSGga | GDFEKCLAtt v~                |                              |
| <i>T. pubescens</i>                 | GvCtLDAFVE SqAYARNDge | GDFEKCFAt~ ~                 |                              |
| <i>A. pediades</i>                  | SlCtLEAFVE SqkYAReDgq | GDFEKCFD~ ~                  |                              |
| <i>P. lycii</i>                     | GvCELsAFVE SqTYAReNgq | GDFAKCgfv se                 |                              |
| <i>A. terreus</i> 9a1               | GRCKrDAFVA GLSFAQAG.. | GNWADCF~~~                   |                              |
| <i>A. terreus</i> cbs               | GRCKrDDFVE GLSFARAG.. | GNWAECF~~~                   |                              |
| <i>A. niger</i> var. <i>awamori</i> | GRCtrDsFVr GLSFARSG.. | GDWAECsA~~                   |                              |
| <i>A. niger</i> T213                | GRCtrDsFVr GLSFARSG.. | GDWAECFA~~                   |                              |
| <i>A. niger</i> NRRL3135            | GRCtrDsFVr GLSFARSG.. | GDWAECFA~~                   |                              |
| <i>A. fumigatus</i> ATCC13073       | GRCKLNDFVK GLSWARSG.. | GNWGEcFS~~                   |                              |
| <i>A. fumigatus</i> ATCC32722       | GRCKLNDFVK GLSWARSG.. | GNWGEcFS~~                   |                              |
| <i>A. fumigatus</i> ATCC58128       | GRCKLNDFVK GLSWARSG.. | GNWGEcFS~~                   |                              |
| <i>A. fumigatus</i> ATCC26906       | GRCKLNDFVK GLSWARSG.. | GNWGEcFS~~                   |                              |
| <i>A. fumigatus</i> ATCC32239       | GRCKLKDFVK GLSWARSG.. | GNSEQSFS~~                   |                              |
| <i>E. nidulans</i>                  | GRCTLDDWVE GLNFARSG.. | GNWktCFTl~                   |                              |
| <i>T. thermophilus</i>              | GRCKrDDFVr GLSFARqG.. | GNWEGCYAas e~                |                              |
| <i>T. lanuginosus</i>               | GRCRrDEWIK GLTFARqG.. | GHWDrcF~~~                   |                              |
| <i>M. thermophila</i>               | GmCtLErFIE SMAFARGN.. | GKWDlCFA~~                   |                              |
| <b>Consensus Seq. 11</b>            | <b>GRCKLDDFVE</b>     | <b>GLSFARSG-- GNWAECFA--</b> | <b>--</b>                    |

Fig. 6e

M G V F V V L L S I A T L F G S T S G T 20  
 ATGGGCGTGTTCGTCTGCTACTGTCCATTGCCACCTTGTTTCGGTTCCACATCCGGTACC  
 1 -----+-----+-----+-----+-----+-----+ 60  
 TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG  
  
 A L G P R G N S H S C D T V D G G Y Q C 40  
 GCCTTGGGTCCTCGTGGTAATTCTCACTCTTG TGACACTGTTGACGGTGGTTACCAATGT  
 61 -----+-----+-----+-----+-----+-----+ 120  
 CGGAACCCAGGAGCACCATTAAGAGTGAGAACTGTGACAAC TGCCACCAATGGTTACA  
  
 F P E I S H L W G T Y S P Y F S L A D E 60  
 TTCCAGAAATTTCTCACTTGTTGGGGTACCTACTCTCCATACTTCTCTTTGGCAGACGAA  
 121 -----+-----+-----+-----+-----+-----+ 180  
 AAGGGTCTTTAAAGAGTGAACACCCCATGGATGAGAGGTATGAAGAGAAACCGTCTGCTT  
  
 S A I S P D V P D D C R V T F V Q V L S 80  
 TCTGCTATTTCTCCAGACGTTCCAGACGACTGTAGAGTTACTTTTCGTTCAAGTTTTGTCT  
 187 -----+-----+-----+-----+-----+-----+ 240  
 AGACGATAAAGAGGTCTGCAAGGTCTGCTGACATCTCAATGAAAGCAAGTTCAAAACAGA  
  
 R H G A R Y P T S S A S K A Y S A L I E 100  
 AGACACGGTGCTAGATACCCAAC TTCTTCTGCGTCTAAGGCTTACTCTGCTTTGATTGAA  
 241 -----+-----+-----+-----+-----+-----+ 300  
 TCTGTGCCACGATCTATGGGTTGAAGAAGACGAGATTCCGAATGAGACGAAACTAACTT  
  
 A I Q K N A T A F K G K Y A F L K T Y N 120  
 GCTATTCAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAC  
 301 -----+-----+-----+-----+-----+-----+ 360  
 CGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTG  
  
 Y T L G A D D L T P F G E N Q M V N S G 140  
 TACACTTTGGGTGCTGACGACTTGACTCCATTTCGGTGAAAACCAAATGGTTAACTCTGGT  
 361 -----+-----+-----+-----+-----+-----+ 420  
 ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACCAATTGAGACCA  
  
 I K F Y R R Y K A L A R K I V P F I R A 160  
 ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCATTCAATTAGAGCT  
 421 -----+-----+-----+-----+-----+-----+ 480  
 TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA  
  
 S G S D R V I A S A E K F I E G F Q S A 180  
 TCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTTTCCAATCTGCT  
 481 -----+-----+-----+-----+-----+-----+ 540  
 AGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAAC TTCAAAGGTTAGACGA  
  
 K L A D P G S Q P H Q A S P V I N V I I 200  
 AAGTTGGCTGACCCAGGTTCTCAACCACACCAAGCTTCTCCAGTTATTAACGTGATCATT  
 541 -----+-----+-----+-----+-----+-----+ 600  
 TTCAACCGACTGGGTCCAAGAGTTGGTGTGGTTTCAAGAGGTCAATAATTGCACTAGTAA  
  
 P E G S G Y N N T L D H G T C T A F E D 220  
 CCAGAAGGATCCGGTTACAACAACACTTTGGACCACGGTACTTGTACTGCTTTTGAAGAC  
 601 -----+-----+-----+-----+-----+-----+ 660  
 GGTCTTCTAGGCCAATGTTGTTGTGAAACCTGGTGCCATGAACATGACGAAAGCTTCTG

Fig. 7a

S E L G D D V E A N F T A L F A P A I R 240  
 TCTGAATTAGGTGACGACGTTGAAGCTAACTTCACTGCTTTGTTGCTCCAGCTATTAGA  
 661 -----+-----+-----+-----+-----+-----+ 720  
 AGACTTAATCCACTGCTGCAACTTCGATTGAAGTGACGAAACAAGCGAGGTCGATAATCT  
  
 A R L E A D L P G V T L T D E D V V Y L 260  
 GCTAGATTGGAAGCTGACTTGCCAGGTGTTACTTTGACTGACGAAGACGTTGTTTACTTG  
 721 -----+-----+-----+-----+-----+-----+ 780  
 CGATCTAACCTTCGACTGAACGGTCCACAATGAAACTGACTGCTTCTGCAACAAATGAAC  
  
 M D M C P F D T V A R T S D A T E L S P 280  
 ATGGACATGTGTCCATTTCGACACTGTCGCTAGAAGTCTGACGCTACTGAATTGTCTCCA  
 781 -----+-----+-----+-----+-----+-----+ 840  
 TACCTGTACACAGGTAAGCTGTGACAGCGATCTTGAAGACTGCGATGACTTAACAGAGGT  
  
 F C A L F T H D E W I Q Y D Y L Q S L G 300  
 TTCTGTGCTTTGTTCACTCACGACGAATGGATCCAATACGACTACTTGCAAAGCTTGGGT  
 841 -----+-----+-----+-----+-----+-----+ 900  
 AAGACACGAAACAAGTGAGTGCTGCTTACCTAGGTTATGCTGATGAACGTTTTCGAACCCA  
  
 K Y Y G Y G A G N P L G P A Q G V G F A 320  
 AAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTGTTGGTTTCGCT  
 901 -----+-----+-----+-----+-----+-----+ 960  
 TTCATGATGCCAATGCCACGACCATTGGGTAAACCCAGGTCGAGTTCACAACCAAAGCGA  
  
 N E L I A R L T H S P V Q D H T S T N H 340  
 AACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTTCTACTAACCAC  
 961 -----+-----+-----+-----+-----+-----+ 1020  
 TTGCTTAACCTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGTAAGATGATTGGTG  
  
 T L D S N P A T F P L N A T L Y A D F S 360  
 ACTTTGGACTCTAACCCAGCTACTTTCCCATTTGAACGCTACTTTGTACGCTGACTTCTCT  
 1021 -----+-----+-----+-----+-----+-----+ 1080  
 TGAAACCTGAGATTGGGTGCTGATGAAAGGGTAACTTGCGATGAAACATGCGACTGAAGAGA  
  
 H D N T M I S I F F A L G L Y N G T K P 380  
 CACGACAACACTATGATATCTATTTCTTCGCTTTGGGTTTGTACAACGGTACCAAGCCA  
 1081 -----+-----+-----+-----+-----+-----+ 1140  
 GTGCTGTTGTGATACTATAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGGTTCCGT  
  
 L S T T S V E S I E E T D G Y S A S W T 400  
 TTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACTCTGCTTCTTGGACT  
 1141 -----+-----+-----+-----+-----+-----+ 1200  
 AACAGATGATGAAGACAACCTTAGATAACTTCTTTGACTGCCAATGAGACGAAGAACCTGA  
  
 V P F A A R A Y V E M M Q C Q A E K E P 420  
 GTTCCATTGCTGCTAGAGCTTACGTTGAAATGATGCAATGTCAAGCTGAAAAGGAACCA  
 1201 -----+-----+-----+-----+-----+-----+ 1260  
 CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACAGTTCGACTTTTCCTTGGT  
  
 L V R V L V N D R V V P L H G C A V D K 440  
 TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGCTGTTGACAAG  
 1261 -----+-----+-----+-----+-----+-----+ 1320  
 AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACGACAACCTGTT

Fig. 7b

24/56

```

L G R C K R D D F V E G L S F A R S G G 460
TTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTAGATCTGGTGGT
1321 -----+-----+-----+-----+-----+ 1380
AACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGATCTAGACCACCA

N W A E C F A * 467
AACTGGGCTGAATGTTTCGCTTAA
1381 -----+-----+----- 1410
TTGACCCGACTTACAAAGCGAATT
```

094822E.0400  
000000 "0928460

Fig. 7c

M G V F V V L L S I A T L F G S T S G T 20  
 ATGGGCGTGTTCGTTCGTGCTACTGTCCATTGCCACCTTGTTCGGTTCACATCCGGTACC  
 1 -----+-----+-----+-----+-----+-----+ 60  
 TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG  
  
 A L G P R G N S H S C D T V D G G Y Q C 40  
 GCCTTGGGTCCCTCGTGGTAACTCTCACTCTTGTGACACTGTTGACGGTGGTTACCAATGT  
 61 -----+-----+-----+-----+-----+-----+ 120  
 CGGAACCCAGGAGCACCATTGAGAGTGAGAACAAGTGTGACAAGTGGCAATGGTTACA  
  
 F P E I S H L W G T Y S P F F S L A D E 60  
 TTCCAGAAATTTCTCACTTGTGGGTACATACTCTCCATTCTTCTCTTTGGCTGACGAA  
 121 -----+-----+-----+-----+-----+-----+ 180  
 AAGGGTCTTTAAAGAGTGAACACCCCATGTATGAGAGGTAAGAAGAGAAACCGACTGCTT  
  
 S A I S P D V P K G C R V T F V Q V L S 80  
 TCTGCTATTTCTCCAGACGTTCCAAAGGGTGTAGAGTTACTTTTCGTTCAAGTTTTGTCT  
 181 -----+-----+-----+-----+-----+-----+ 240  
 AGACGATAAAGAGGTCTGCAAGGTTTCCCAACATCTCAATGAAAGCAAGTTCAAAACAGA  
  
 R H G A R Y P T S S A S K A Y S A L I E 100  
 AGACACGGTGTCTAGATACCCAACTTCTTCTGCGTCTAAGGCGTACTCTGCTTTGATTGAA  
 241 -----+-----+-----+-----+-----+-----+ 300  
 TCTGTGCCACGATCTATGGGTTGAAGAAGACGCAGATTCCGCATGAGACGAAACTAAGTT  
  
 A I Q K N A T A F K G K Y A F L K T Y N 120  
 GCTATTCAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAC  
 301 -----+-----+-----+-----+-----+-----+ 360  
 CGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTG  
  
 Y T L G A D D L T P F G E Q Q M V N S G 140  
 TACACTTTGGGTGCTGACGACTTGACTCCATTTCGGTGAACAACAAATGGTTAACTCTGGT  
 361 -----+-----+-----+-----+-----+-----+ 420  
 ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTGTGTGTTACCAATTGAGACCA  
  
 I K F Y R R Y K A L A R K I V P F I R A 160  
 ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAAGATTGTTCCATTCAATGAGCT  
 421 -----+-----+-----+-----+-----+-----+ 480  
 TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA  
  
 S G S D R V I A S A E K F I E G F Q S A 180  
 TCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTTTCCAATCTGCT  
 481 -----+-----+-----+-----+-----+-----+ 540  
 AGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACTTCCAAAGGTTAGACGA  
  
 K L A D P G A N P H Q A S P V I N V I I 200  
 AAGTTGGCTGACCCAGGTGCTAACCCACACCAAGCTTCTCCAGTTATTAACGTTATTATT  
 541 -----+-----+-----+-----+-----+-----+ 600  
 TTCAACCGACTGGGTCCACGATTGGGTGTGGTTCGAAGAGGTCAATAATTGCAATAATAA  
  
 P E G A G Y N N T L D H G L C T A F E E 220  
 CCAGAAGGTGCTGGTTACAACAACACTTTGGACCACGGTTTGTGTACTGCTTTTCAAGAA  
 601 -----+-----+-----+-----+-----+-----+ 660  
 GGTCTTCCACGACCAATGTTGTTGTGAAACCTGGTGCCAAACACATGACGAAAGCTTCTT

Fig. 8a

S E L G D D V E A N F T A V F A P P I R 240  
TCTGAATTGGGTGACGACGTTGAAGCTAACTTCACTGCTGTTTTTCGCTCCACCAATTAGA  
661 -----+-----+-----+-----+-----+-----+ 720  
AGACTTAACCCACTGCTGCAACTTCGATTGAAGTGACGACAAAAGCGAGGTGGTTAATCT

A R L E A H L P G V N L T D E D V V N L 260  
GCTAGATTGGAAGCTCACTTGCCAGGTGTTAACTTGACTGACGAAGACGTTGTTAACTTG  
721 -----+-----+-----+-----+-----+-----+ 780  
CGATCTAACCTTCGAGTGAACGGTCCACAATTGAACTGACTGCTTCTGCAACAATTGAAC

M D M C P F D T V A R T S D A T Q L S P 280  
ATGGACATGTGTCCATTCGACACTGTTGCTAGAACTTCTGACGCTACTCAATTGTCTCCA  
781 -----+-----+-----+-----+-----+-----+ 840  
TACCTGTACACAGGTAAGCTGTGACAACGATCTTGAAGACTGCGATGAGTTAACAGAGGT

F C D L F T H D E W I Q Y D Y L Q S L G 300  
TTCTGTGACTTGTTCACCTCACGACGAATGGATTCAATACGACTACTTGCAATCTTTGGGT  
841 -----+-----+-----+-----+-----+-----+ 900  
AAGACACTGAACAAGTGAGTGCTGCTTACCTAAGTTATGCTGATGAACGTTAGAAACCCA

K Y Y G Y G A G N P L G P A Q G V G F V 320  
AAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTGTTGGTTTCGTT  
901 -----+-----+-----+-----+-----+-----+ 960  
TTCATGATGCCAATGCCACGACCATTGGGTAAACCCAGGTCGAGTTCCACAACCAAAGCAA

N E L I A R L T H S P V Q D H T S T N H 340  
AACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTTCTACTAACCAC  
961 -----+-----+-----+-----+-----+-----+ 1020  
TTGCTTAACTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGGAAGATGATTGGTG

T L D S N P A T F P L N A T L Y A D F S 360  
ACTTTGGACTCTAACCCAGCTACTTTCCCATGAAACGCTACTTTGTACGCTGACTTCTCT  
1021 -----+-----+-----+-----+-----+-----+ 1080  
TGAAACCTGAGATTGGGTGCGATGAAAGGGTAACTTGCGATGAAACATGCGACTGAAGAGA

H D N T M V S I F F A L G L Y N G T K P 380  
CACGACAACACTATGTTTCTATTTTCTTCGCTTTGGGTTTGTACAACGGTACTAAGCCA  
1081 -----+-----+-----+-----+-----+-----+ 1140  
GTGCTGTTGTGATACCAAAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGATTCCGGT

L S T T S V E S I E E T D G Y S A S W T 400  
TTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACTCTGCTTCTTGGACT  
1141 -----+-----+-----+-----+-----+-----+ 1200  
AACAGATGATGAAGACAACCTTAGATAACTTCTTTGACTGCCAATGAGACGAAGAACCTGA

V P F A A R A Y V E M M Q C E A E K E P 420  
GTTCCATTGCTGCTAGAGCTTACGTTGAAATGATGCAATGTGAAGCTGAAAAGGAACCA  
1201 -----+-----+-----+-----+-----+-----+ 1260  
CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACACTTCGACTTTTCCTTGGT

L V R V L V N D R V V P L H G C G V D K 440  
TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGGTGTTGACAAG  
1261 -----+-----+-----+-----+-----+-----+ 1320  
AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACCACAACCTGTTCT

Fig. 8b

[illegible]

**Figure 1**

[illegible]

M G V F V V L L S I A T L F G S T S G T 20  
 ATGGGGGTTTTCGTCGTTCTATTATCTATCGCGACTCTGTTCGGCAGCACATCGGGCACT  
 1 -----+-----+-----+-----+-----+-----+ 60  
 TACCCCCAAAAGCAGCAAGATAATAGATAGCGCTGAGACAAGCCGTCGTGTAGCCCGTGA  
  
 A L G P R G N H S K S C D T V D L G Y Q 40  
 GCGCTGGGCCCCCGTGAAATCACTCCAAGTCTCGGATACGGTAGACCTAGGGTACCAG  
 61 -----+-----+-----+-----+-----+-----+ 120  
 CGCGACCCGGGGGCACCTTTAGTGAGGTTTCAGGACGCTATGCCATCTGGATCCCATGGTC  
  
 C S P A T S H L W G T Y S P Y F S L E D 60  
 TGCTCCCTGCGACTTCTCATCTATGGGGCAGTACTCGCCATaCTTTTCGCTCGAGGAC  
 121 -----+-----+-----+-----+-----+-----+ 180  
 ACGAGGGGACGCTGAAGAGTAGATACCCCGtgCATGAGCGGTAtGAAAAGCGAGCTCCTG  
  
 E L S V S S K L P K D C R I T L V Q V L 80  
 GAGCTGTCCGTGTCGAGTAAGCTTCCCAAGGATTGCCGGATCACCTTGGTACAGGTGCTA  
 181 -----+-----+-----+-----+-----+-----+ 240  
 CTCGACAGGCACAGCTCATTCGAAGGGTTCTAACGGCCTAGTGGAACCATGTCCACGAT  
  
 S R H G A R Y P T S S K S K K Y K K L I 100  
 TCGCGCCATGGAGCGCGGTACCCAACCAGCTCCAAGAGCAAAAAGTATAAGAAGCTTaTt  
 241 -----+-----+-----+-----+-----+-----+ 300  
 AGCGCGGTACCTCGCGCCATGGGTTGGTTCGAGGTTCTCGTTTTTCATATTCTTCGAAtAa  
  
 T A I Q A N A T D F K G K Y A F L K T Y 120  
 ACGGCGATCCAGGCCAATGCCACCGACTTCAAGGGCAAGTAcGCCTTTTTGAAGACGTAC  
 301 -----+-----+-----+-----+-----+-----+ 360  
 TGCCGCTAGGTCCGGTTACGGTGGCTGAAGTTCCCGTTCAtgCGGAAAACTTCTGCATG  
  
 N Y T L G A D D L T P F G E Q Q L V N S 140  
 AACTATACTCTGGGTGCGGATGACCTCACTCCCTTTGGGGAGCAGCAGCTGGTGAACCTCG  
 361 -----+-----+-----+-----+-----+-----+ 420  
 TTGATATGAGACCCACGCTACTGGAGTGAGGGAAACCCCTCGTCGTCGACCACTTGAGC  
  
 G I K F Y Q R Y K A L A R S V V P F I R 160  
 GGCATCAAGTTCTACCAGAGGTACAAGGCTCTGGCGCGCAGTGTGGTGCCGTTTATTTCGC  
 421 -----+-----+-----+-----+-----+-----+ 480  
 CCGTAGTTCAAGATGGTCTCCATGTTCCGAGACCGCGCTCACACCACGGCAAATAAGCG  
  
 A S G S D R V I A S G E K F I E G F Q Q 180  
 GCCTCAGGCTCGGACCGGGTTATTGCTTCGGGAGAGAAGTTCATCGAGGGGTTCCAGCAG  
 481 -----+-----+-----+-----+-----+-----+ 540  
 CGGAGTCCGAGCCTGGCCCAATAACGAAGCCCTCTCTTCAAGTAGCTCCCCAAGGTCGTC  
  
 A K L A D P G A T N R A A P A I S V I I 200  
 GCGAAGCTGGCTGATCCTGGCGCGACGAACCGCGCCGCTCCGGCGATTAGTGTGATTATT  
 541 -----+-----+-----+-----+-----+-----+ 600  
 CGCTTCGACCGACTAGGACCGCGCTGCTTGGCGCGCGAGGCCGCTAATCACACTAATAA  
  
 P E S E T F N N T L D H G V C T K F E A 220  
 CCGGAGAGCGAGACGTTCAACAATACGCTGGACCACGGTGTGTGCACGAAGTTTGAGGCG  
 601 -----+-----+-----+-----+-----+-----+ 660  
 GGCCTCTCGCTCTGCAAGTTGTTATGCGACCTGGTGCCACACACGTGCTTCAAACCTCCGC

Fig. 9a

S Q L G D E V A A N F T A L F A P D I R 240  
 AGTCAGCTGGGAGATGAGGTTGCGGCCAATTTCACTGCGCTCTTTGCACCCGACATCCGA  
 661 -----+-----+-----+-----+-----+-----+ 720  
 TCAGTCGACCCTCTACTCCAACGCCGGTTAAAGTGACGCGAGAAACGTGGGCTGTAGGCT  
  
 A R L E K H L P G V T L T D E D V V S L 260  
 GCTCGCctCGAGAAGCATCTTCCTGGCGTGACGCTGACAGACGAGGACGTTGTCAGTCTA  
 721 -----+-----+-----+-----+-----+-----+ 780  
 CGAGCGgaGCTCTTCGTAGAAAGGACCGCACTGCGACTGTCTGCTCCTGCAACAGTCAGAT  
  
 M D M C P F D T V A R T S D A S Q L S P 280  
 ATGGACATGTGTcCGTTTGATACGGTAGCGCGCACCAGCGACGCAAGTCAGCTGTCACCG  
 781 -----+-----+-----+-----+-----+-----+ 840  
 TACCTGTACACAgGCAAACATATGCCATCGCGCGTGGTCGCTGCGTTTCAGTCGACAGTGGC  
  
 F C Q L F T H N E W K K Y D Y L Q S L G 300  
 TTCTGTCAACTCTTCACTCACAATGAGTGGAAGAAGTACgACTACCTTCAGTCCTTGCGC  
 841 -----+-----+-----+-----+-----+-----+ 900  
 AAGACAGTTGAGAAGTGAGTGTTACTCACCTTCTTCATGcTGATGGAAGTCAGGAACCCG  
  
 K Y Y G Y G A G N P L G P A Q G I G F T 320  
 AAGTACTACGGCTACGGCGCAGGCAACCCTCTGGGACCGGCTCAGGGGATAGGGTTACCC  
 901 -----+-----+-----+-----+-----+-----+ 960  
 TTCATGATGCCGATGCCCGCTCCGTTGGGAGACCCTGGCCGAGTCCCCTATCCCAAGTGG  
  
 N E L I A R L T R S P V Q D H T S T N S 340  
 AACGAGCTGATTGCCCGGTTGACgCGTTGCCAGTGACAGGACCACACCAGCACTAACTCG  
 961 -----+-----+-----+-----+-----+-----+ 1020  
 TTGCTCGACTAACGGGCCAACTGcGCAAGCGGTACAGTCCTGGTGTGGTCGTGATTGAGC  
  
 T L V S N P A T F P L N A T M Y V D F S 360  
 ACTCTAGTCTCCAACCCGGCCACCTTCCCCGTTGAACGCTACCATGTACGTCGACTTTTCA  
 1021 -----+-----+-----+-----+-----+-----+ 1080  
 TGAGATCAGAGGTTGGGCCCGTGAAGGGCAACTTGCGATGGTACATGCAGCTGAAAAGT  
  
 H D N S M V S I F F A L G L Y N G T E P 380  
 CACGACAACAGCATGGTTTCCATCTTCTTTGCATTGGGCCTGTACAACGGCACTGAACCC  
 1081 -----+-----+-----+-----+-----+-----+ 1140  
 GTGCTGTTGTCTGTACCAAAGGTAGAAGAAACGTAACCCGGACATGTTGCCGTGACTTGGG  
  
 L S R T S V E S A K E L D G Y S A S W V 400  
 TTGTCCCGGACCTCGGTGGAAAGCGCCAAGGAATTGGATGGGTATTCTGCATCCTGGGTG  
 1141 -----+-----+-----+-----+-----+-----+ 1200  
 AACAGGGCCTGGAGCCACCTTTCGCGGTTCTTAACCTACCCATAAGACGTAGGACCCAC  
  
 V P F G A R A Y F E T M Q C K S E K E P 420  
 GTGCCTTTTCGGCGCGCGAGCCTACTTCGAGACGATGCAATGCAAGTCGGAAGAGGAGCCT  
 1201 -----+-----+-----+-----+-----+-----+ 1260  
 CACGGAAAGCCGCGCGCTCGGATGAAGCTCTGCTACGTTACGTTACGCTTTTCTCGGA  
  
 L V R A L I N D R V V P L H G C D V D K 440  
 CTTGTTTCGCGCTTTGATTAATGACCGGGTTGTGCCACTGCATGGCTGCGATGTGGACAAG  
 1261 -----+-----+-----+-----+-----+-----+ 1320  
 GAACAAGCGCGAAACTAATTACTGGCCCAACACGGTGACGTACCGACGCTACACCTGTTT

Fig. 9b

30/56

```

 L G R C K L N D F V K G L S W A R S G G 460
 CTGGGGCGATGCAAGCTGAATGACTTTGTCAAGGGATTGAGTTGGGCCAGATCTGGGGGC
1321 -----+-----+-----+-----+-----+ 1380
 GACCCCGCTACGTTGACTTACTGAAACAGTTCCTAACTCAACCCGGTCTAGACCCCG

 N W G E C F S * 467
 AACTGGGGAGAGTGCTTTAGTTGA
1381 -----+-----+----- 1404
 TTGACCCCTCTCACGAAATCAACT
```

Fig. 9c

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted March 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CP-1  
*Eco* RI M G V F V V L L S I A T L F G S T  
**TATATGAATTCATGGGCGTGTTCGTGCTACTGTCCATTGCCACCTTGTTCGGTTCCA**  
1 -----+-----+-----+-----+-----+-----+ 60  
ATATACTTAAGTACCCGCACAAGCAGCACGATGACAGGTAACGGTGAACAAGCCAAGGT  
S G T A L G P R G N S H S C D T V D G G  
**CATCCGGTACCGCCTTGGGTCTCGTGGTAATTCTCACTCTTGTGACACTGTTGACGGTG**  
61 -----+-----+-----+-----+-----+-----+ 120  
GTAGGCCATGGCGGAACCCAGGAGCACCATTAAAGAGTGAGAACACTGTGACAACCTGCCAC  
CP-2  
CP-3  
Y Q C F P E I S H L W G Q Y S P Y F S L  
GTTACCAATGTTTCCAGAAATTTCTCACTTGTGGGGTCAATACTCTCCATACTTCTCTTT  
121 -----+-----+-----+-----+-----+-----+ 180  
CAATGGTTACAAAGGGTCTTTAAAGAGTGAACACCCCAGTTATGAGAGGTATGAAGAGAA  
E D E S A I S P D V P D D C R V T F V Q  
**TGGAAGACGAATCTGCTATTTCTCCAGACGTTCCAGACGACTGTAGAGTTACTTTCGTTTC**  
181 -----+-----+-----+-----+-----+-----+ 240  
ACCTTCTGCTTAGACGATAAAGAGGTCTGCAAGGTCTGCTGACATCTCAATGAAAGCAAG  
CP-4.7  
CP-5.7  
V L S R H G A R Y P T D S K G K K Y S A  
AAGTTTGTCTAGACACGGTGCTAGATACCCAAGTgacTCTAAGggtAAGaagTACTCTG  
241 -----+-----+-----+-----+-----+-----+ 300  
TTCAAAACAGATCTGTGCCACGATCTATGGGTGActgAGATTCCaTTCTtcATGAGAC  
L I E A I Q K N A T A F K G K Y A F L K  
**CTTTGATTGAAGCTATTCAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGA**  
301 -----+-----+-----+-----+-----+-----+ 360  
GAAACTAACTTCGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAAGT  
CP-6  
CP-7  
T Y N Y T L G A D D L T P F G E N Q M V  
AGACTTACAACCTACACTTTGGGTGCTGACGACTTGACTCCATTCCGGTGAAAACCAAATGG  
361 -----+-----+-----+-----+-----+-----+ 420  
TCTGAATGTTGATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACC  
N S G I K F Y R R Y K A L A R K I V P F  
**TAACTCTGGTATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCAT**  
421 -----+-----+-----+-----+-----+-----+ 480  
AATTGAGACCATAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTA  
CP-8.7  
CP-9  
I R A S G S S R V I A S A E K F I E G F  
TCATTAGAGCTTCTGGTTCTTctAGAGTTATTGCTTCTGCTGAAAAGTTTCATTGAAGGTT  
481 -----+-----+-----+-----+-----+-----+ 540  
AGTAATCTCGAAGACCAAGAgaTCTCAATAACGAAGACGACTTTTCAAGTAACCTCCAA  
Q S A K L A D P G S Q P H Q A S P V I D  
**TCCAATCTGCTAAGTTGGCTGACCCAGGTTCTCAACCACACCAAGCTTCTCCAGTTATTG**  
541 -----+-----+-----+-----+-----+-----+ 600  
AGGTTAGACGATTCAACCGACTGGGTCCAAGAGTTGGTGTGGTTTCAAGAGGTCAATAAC

Fig. 10a

## CP-10.7

## CP-11.7

V I I S E A S S Y N N T L D P G T C T A  
 ACGTTATTATTtctGACgctTCTtctTACAACAACACTTTGGACccaGGTACTTGTACTG  
 601 -----+-----+-----+-----+-----+ 660  
 TGCAATAATAAagaCTgCGaAGGagaATGTTTGTGTGAAACCTGggtCCATGAACATGAC

F E D S E L A D T V E A N F T A L F A P  
 CTTTCGAAGACTCTGAATTGgctGACactGTTGAAGCTAACTTCACTGCTTTGTTGCTC  
 661 -----+-----+-----+-----+-----+ 720  
 GAAAGCTTCTGAGACTTAACcgaCTGtgaCAACTTCGATTGAAGTGACGAAACAAGCGAG

## CP-12.7

A I R A R L E A D L P G V T L T D T E V  
 CAGCTATTAGAGCTAGATTGGAAGCTGACTTGCCAGGTGTTACTTTGACTGACactgaaG  
 721 -----+-----+-----+-----+-----+ 780  
 GTCGATAATCTCGATCTAACCTTCGACTGAACGGTCCACAATGAACTGACTGtgacttc

## CP-13.7

T Y L M D M C S F E T V A R T S D A T E  
 TTactTACTTGATGGACATGTGTtctTTTGAAACTGTTGCTAGAACTTCTGACGCTACTG  
 781 -----+-----+-----+-----+-----+ 840  
 AatgaATGAACTACCTGTACACAagaAAGCTTTGACAACGATCTTGAAGACTGCGATGAC

L S P F C A L F T H D E W R H Y D Y L Q  
 AATTGTCTCCATTCTGTGCTTTGTTCACTCACGACGAATGGAGAcacTACGACTACTTGC  
 841 -----+-----+-----+-----+-----+ 900  
 TTAACAGAGGTAAGACACGAAACAAGTGAGTGCTGCTTACCTCTgtgATGCTGATGAACG

## CP-14.7

## CP-15.7

S L K K Y Y G H G A G N P L G P T Q G V  
 AATCTTTGaagAAGTACTACGGTcacGGTGCTGGTAACCCATTGGGTCCAactCAAGGTG  
 901 -----+-----+-----+-----+-----+ 960  
 TTAGAAACttcTTCATGATGCCAgTgCCACGACCATTGGGTAACCCAGGTtgaGTTCCAC

G F A N E L I A R L T R S P V Q D H T S  
 TTGGTTTCGCTAACGAATTGATTGCTAGATTGACTAGATCTCCAGTTCAAGACCACACTT  
 961 -----+-----+-----+-----+-----+ 1020  
 AACCAAAGCGATTGCTTAACTAACGATCTAACTGATCTAGAGGTCAAGTTCTGGTGTGAA

## CP-16

## CP-17.7

T N H T L D S N P A T F P L N A T L Y A  
 CTACTAACCCACACTTTGGACTCTAACCCAGCTACTTTCCCATTTGAACGCTACTTTGTACG  
 1021 -----+-----+-----+-----+-----+ 1080  
 GATGATTGGTGTGAAACCTGAGATTGGGTGCGATGAAAGGGTAACCTTGCATGAAACATGC

D F S H D N G I I S I F F A L G L Y N G  
 CTGACTTCTCTCACGACAACggtattATTTCTATTTTCTTCGCTTTGGGTTTGTACAACG  
 1081 -----+-----+-----+-----+-----+ 1140  
 GACTGAAGAGAGTGCTGTTGccataaTAAAGATAAAAGAAGCGAAACCCAAACATGTTGC

## CP-18.7

## CP-19.7

T A P L S T T S V E S I E E T D G Y S S  
 GTACTGCTCCATTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACTCTT  
 1141 -----+-----+-----+-----+-----+ 1200  
 CATGACGAGGTAACAGATGATGAAGACAACCTTAGATAACTTCTTTGACTGCCAATGAGAA

Fig. 10b

33/56

A W T V P F A S R A Y V E M M Q C Q A E  
ctgctTGGACTGTTCCATTTCgcttctAGAGCTTACGTTGAAATGATGCAATGTCAAGCTG  
1201 -----+-----+-----+-----+-----+ 1260  
gacgaACCTGACAAGGTAAGcgaagaTCTCGAATGCAACTTTACTACGTTACAGTTTCGAC  
CP-20  
CP-21  
K E P L V R V L V N D R V V P L H G C A  
AAAAGGAACCATTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTG  
1261 -----+-----+-----+-----+-----+ 1320  
TTTTCTTGTTAACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACAC  
V D K L G R C K R D D F V E G L S F A R  
CTGTTGACAAGTTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTA  
1321 -----+-----+-----+-----+-----+ 1380  
GACAACTGTTCAACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGAT  
CP-22  
S G G N W A E C F A \* Eco RI  
GATCTGGTGGTAACTGGGCTGAATGTTTCGCTTAAGAATTCATATA  
1381 -----+-----+-----+-----+----- 1426  
CTAGACCACCATTGACCCGACTTACAAAGCGAATTCTTAAGTATAT

Fig. 10c

000210 " 33232450

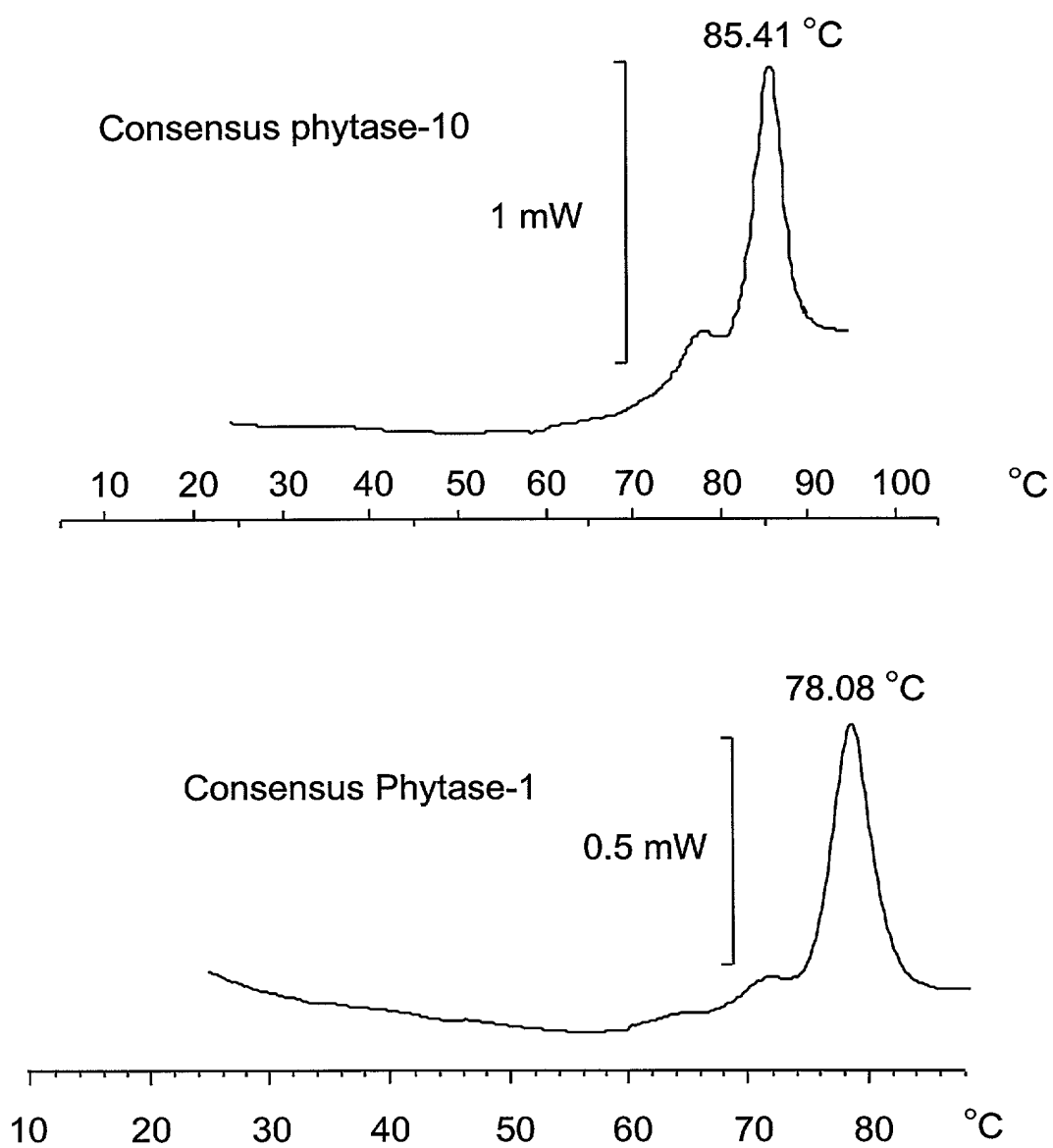


Fig. 11

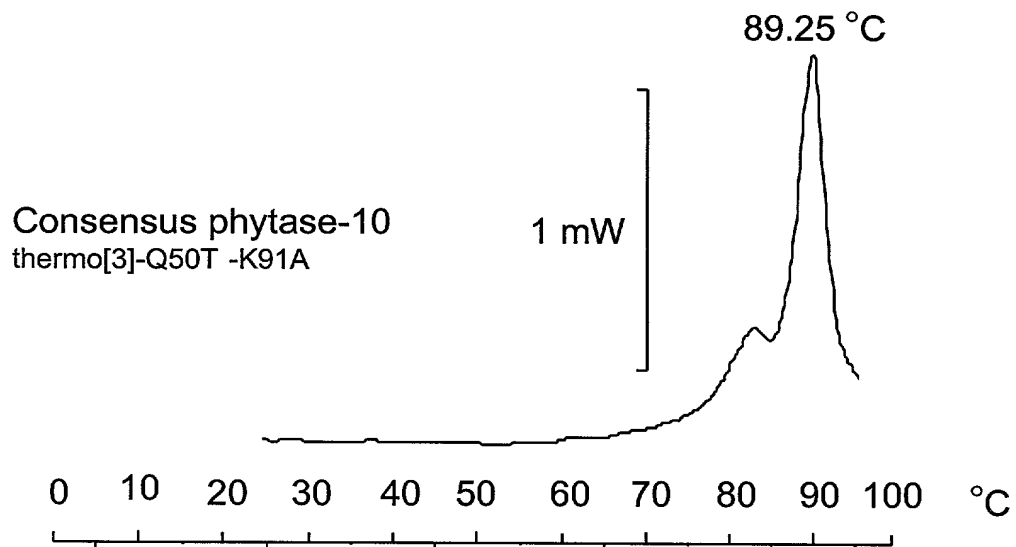
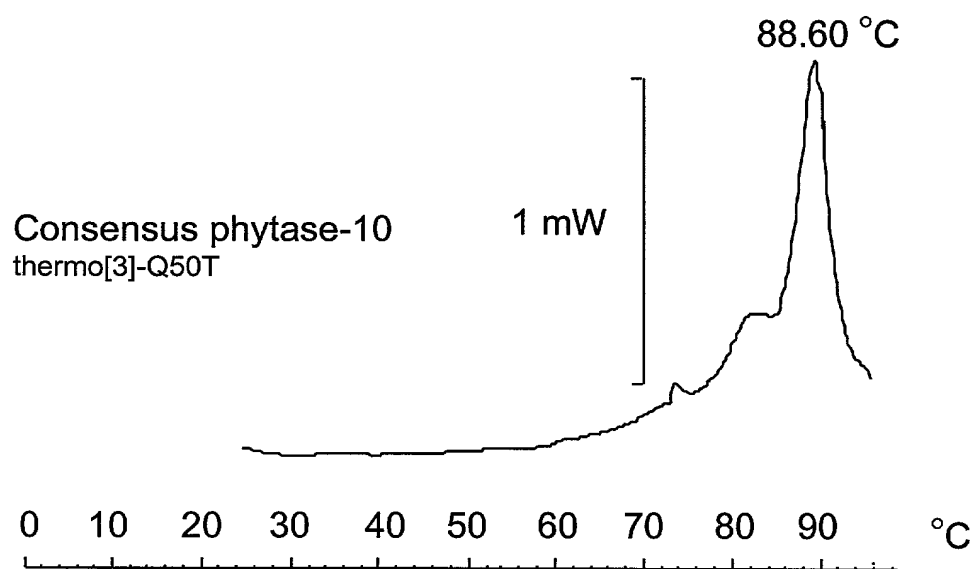


Fig. 12

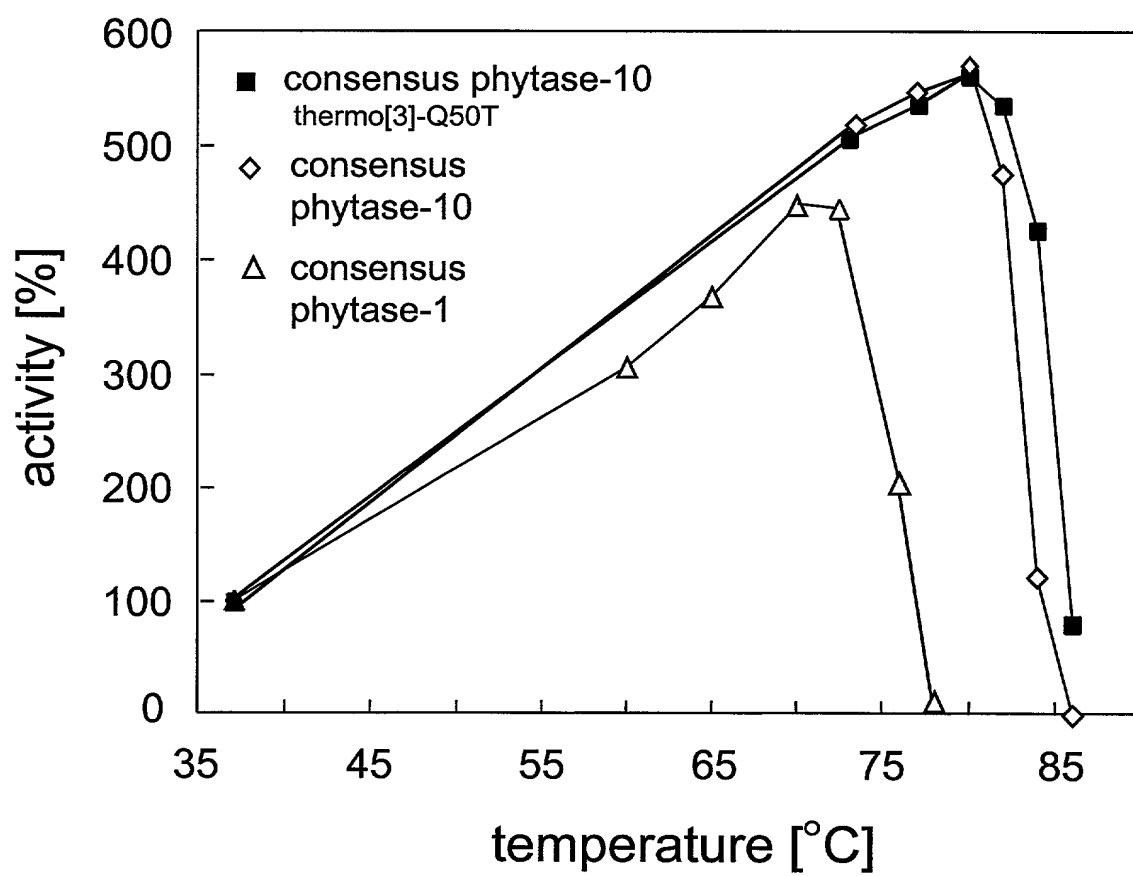


Fig. 13

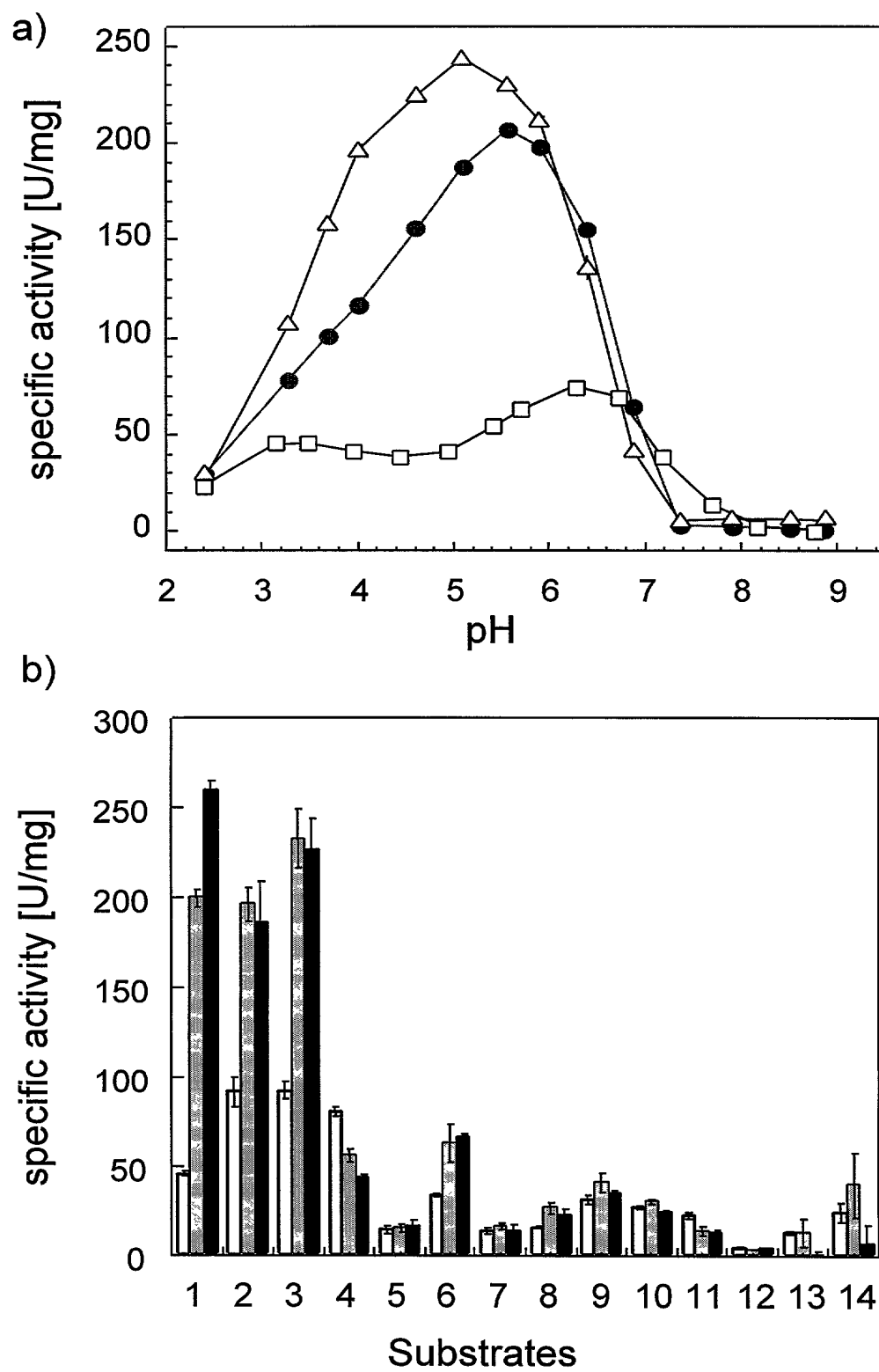


Fig. 14

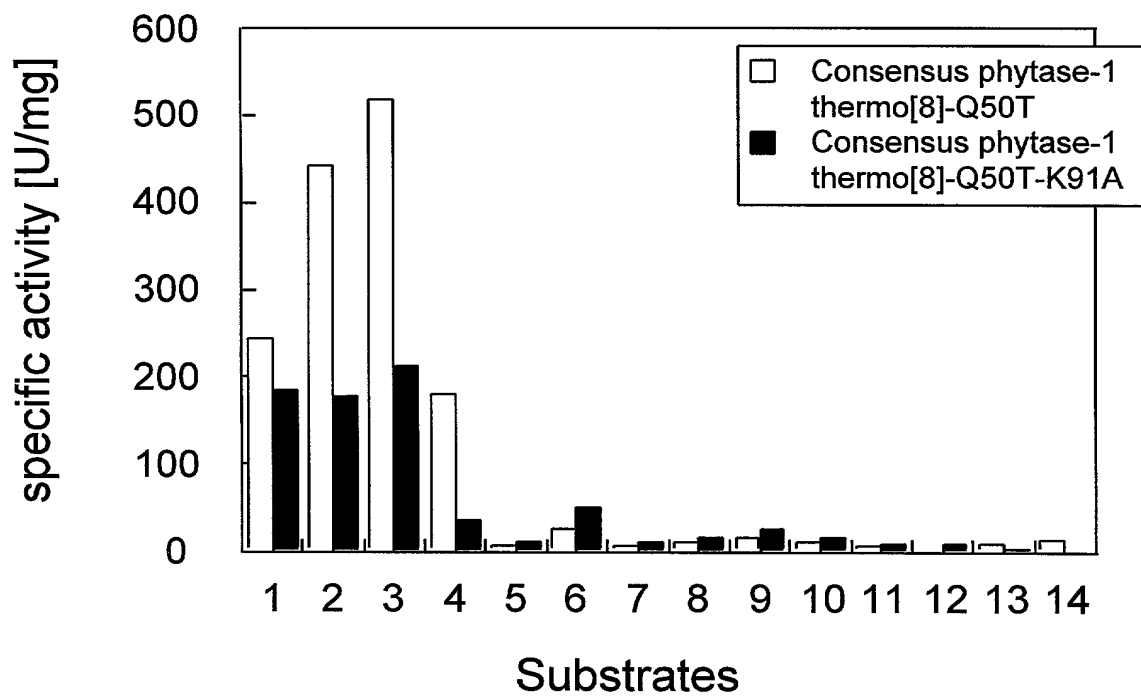
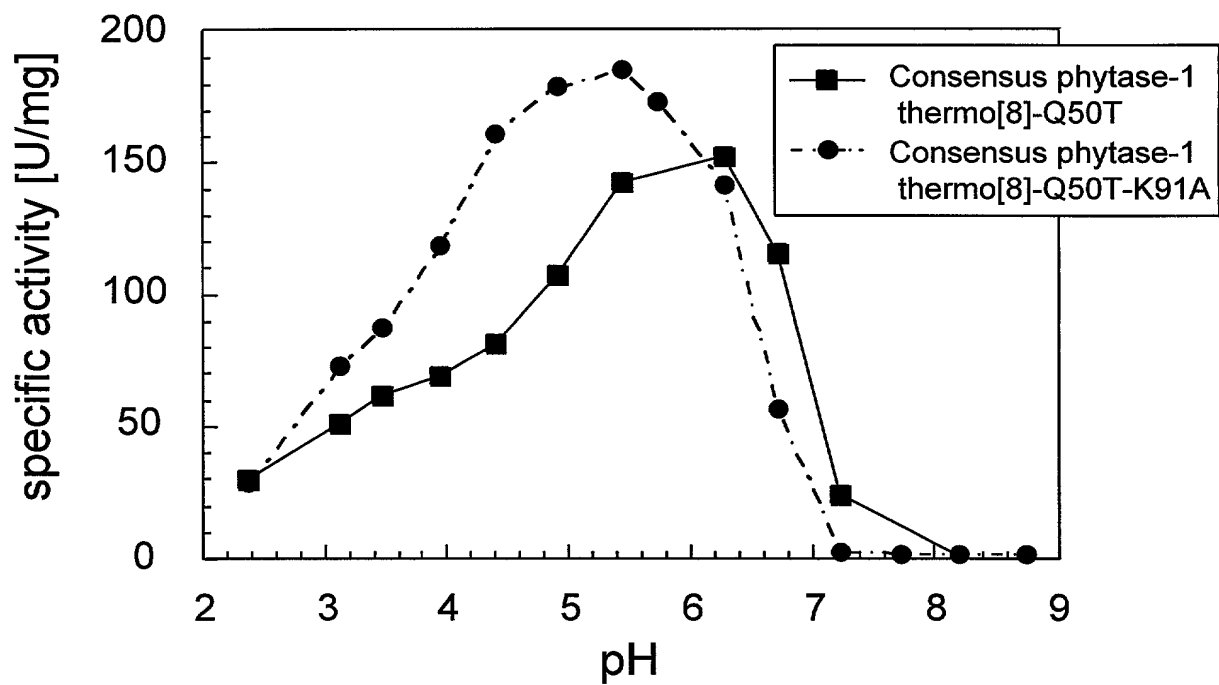


Fig. 15

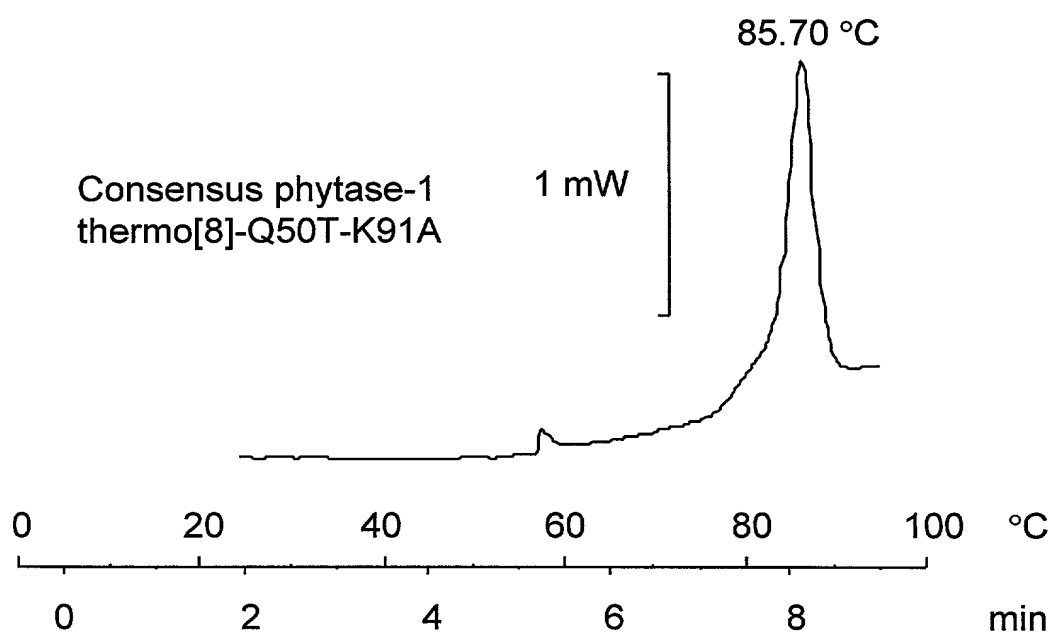
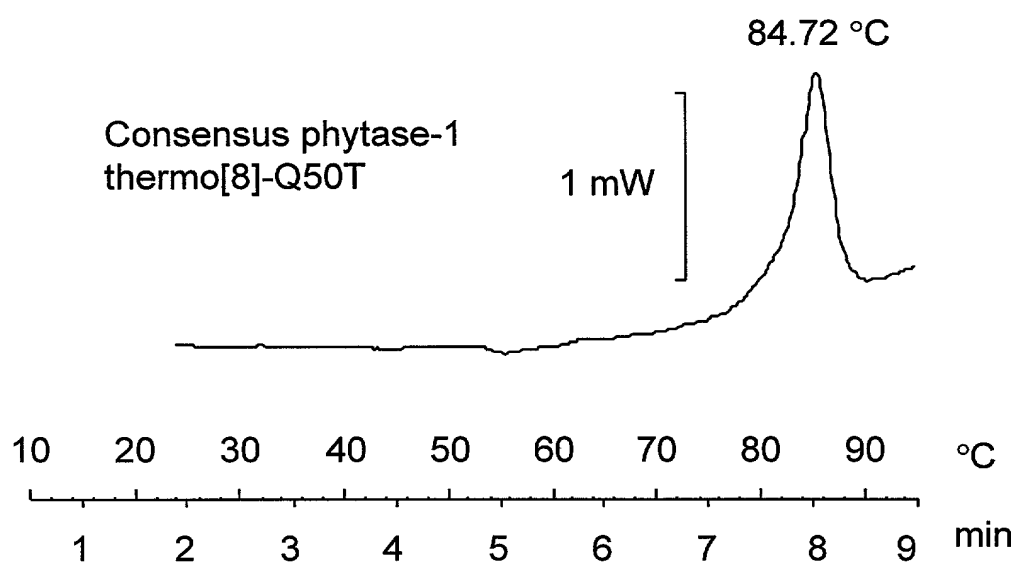


Fig. 16

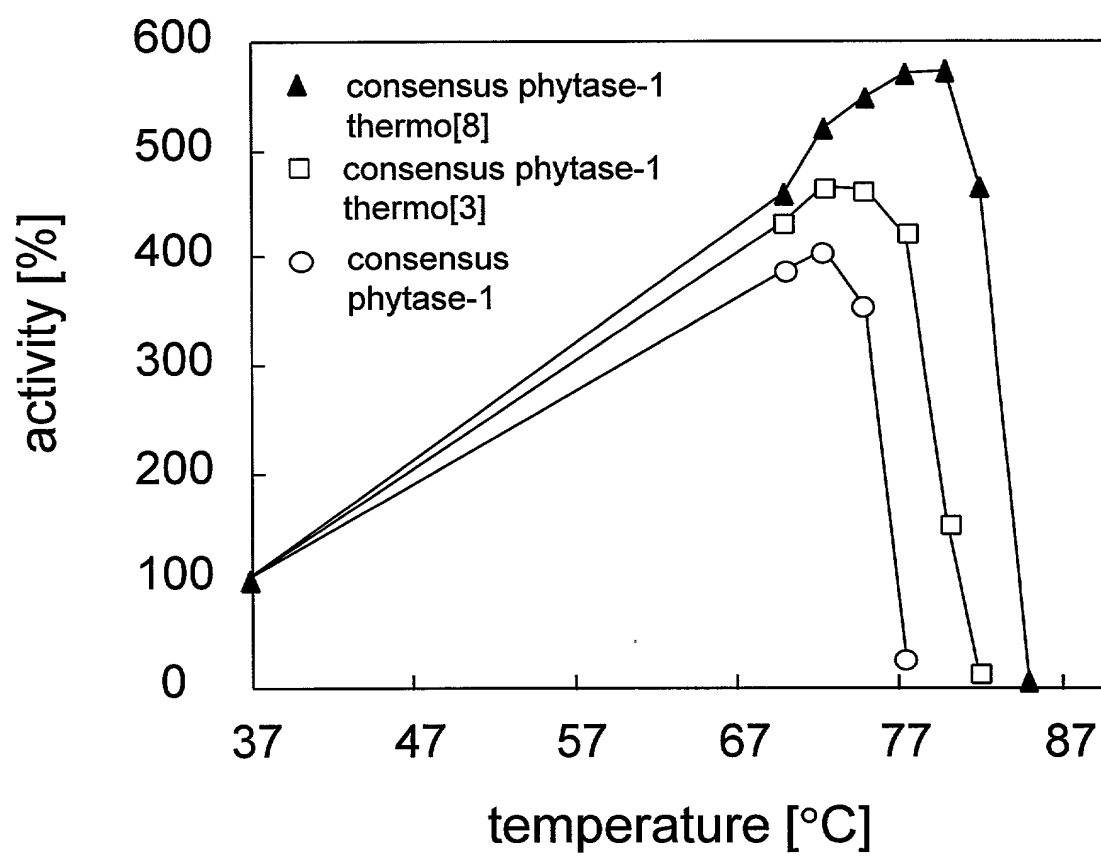


Fig. 17

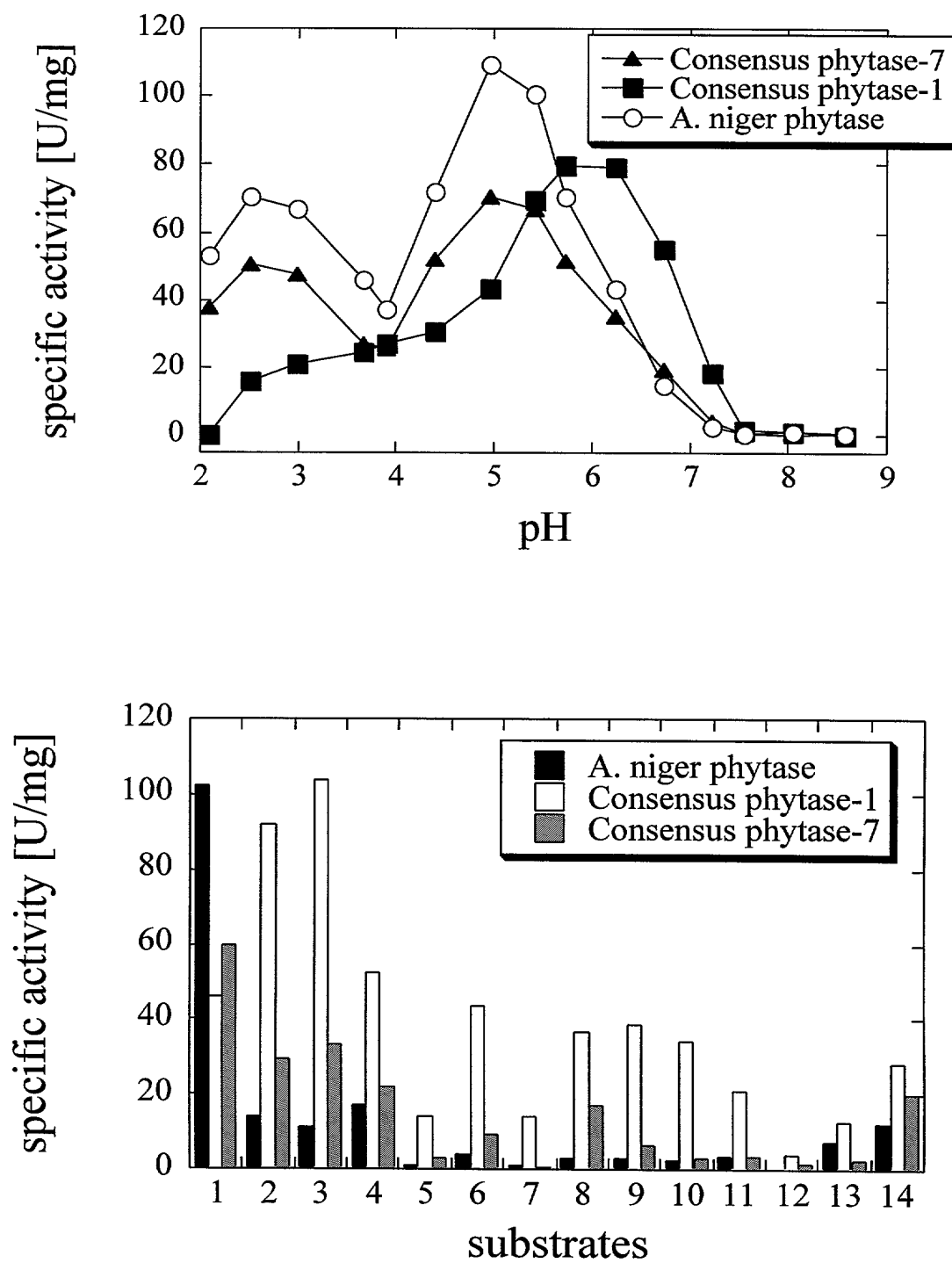


Fig. 18

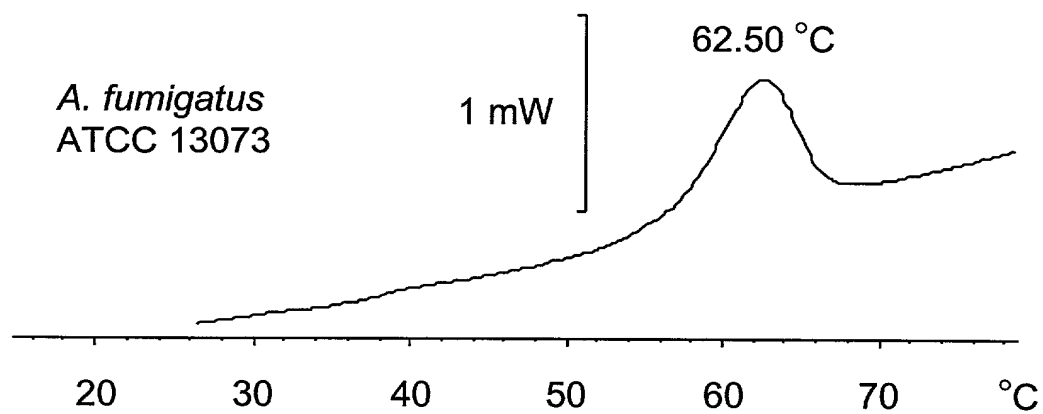
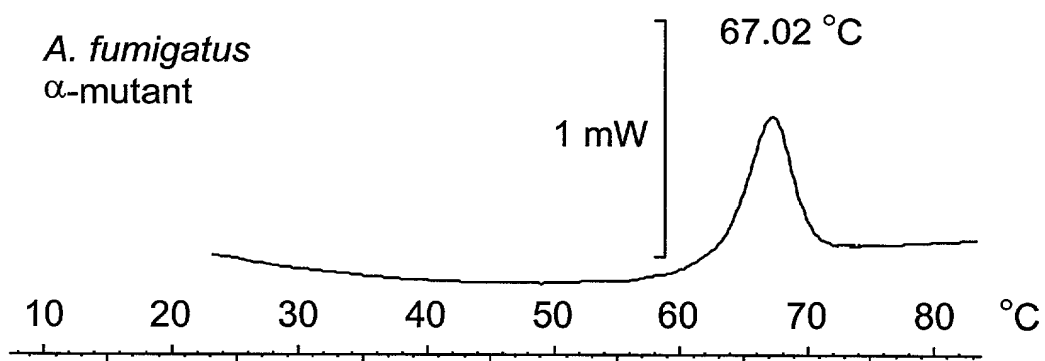


Fig. 19

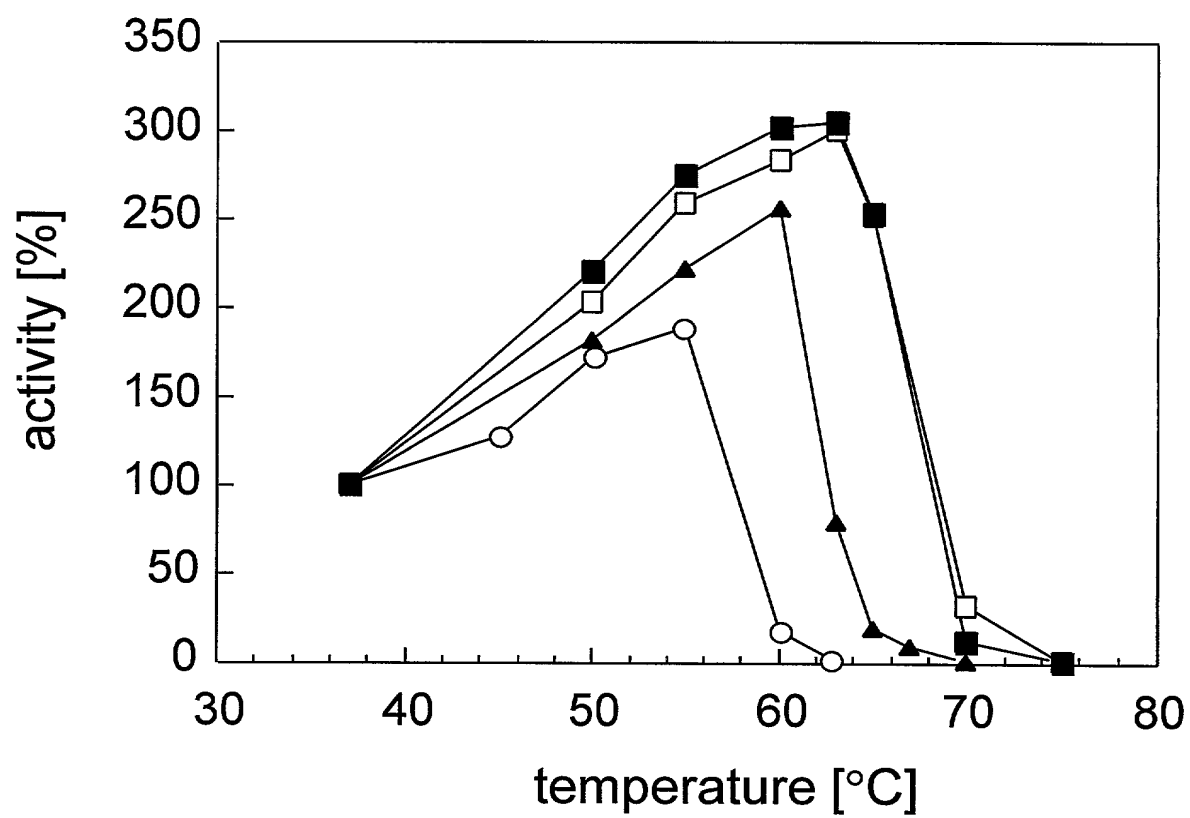


Fig. 20

1     MGVFVLLSI  ATLFGSTSGT  ALGPRGNSHS  CDTVDGGYQC  FPEISSNWSP  
51    YSPYFSLADE  SAISPDVPKG  CRVTFVQVLQ  RHGARFPTSG  AATRISALIE  
101  AIQKNATAFK  GKYAFLKTYN  YTLGADDLVP  FGANQSSQAG  IKFYRRYKAL  
151  ARKIVPFIRA  SGSDRVIDSA  TNWIEGFQSA  KLADPGANPH  QASPVINVII  
201  PEGAGYNNTL  DHGLCTAFEE  SELGDDVEAN  FTAVFAPPIR  ARLEAHLPGV  
251  NLTDDEVVNL  MDMCPFDTVA  RTSDATELSP  FCDLFTHDEW  IQYDYLGDLD  
301  KYYGTGAGNP  LGPAQGVGVFV  NELIARLTHS  PVQDHTSTNH  TLDSNPATFP  
351  LNATLYADFS  HDNTMVAIFF  ALGLYNGTKP  LSTTSVESIE  ETDGYSASWL  
401  VPFSARMYVE  MMQCEAEKEP  LVRVLVNDRV  VPLHGCGVDK  LGRCKRDDFV  
451  EGLSFARSGG  NWEECFA

Fig. 21

ATGGGCGTGTTTCGTGCTACTGTCCATTGCCACCTTGTTTCGGTTCACATCCGGTACC  
 1 -----+-----+-----+-----+-----+-----+ 60  
 TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG  
  
 M G V F V V L L S I A T L F G S T S G T -  
  
 GCCTTGGGTCCTCGTGGTAATTCTCACTCTTGTGACACTGTTGACGGTGGTTACCAATGT  
 61 -----+-----+-----+-----+-----+ 120  
 CGGAACCCAGGAGCACCATTAAGAGTGAGAACACTGTGACAACTGCCACCAATGGTTACA  
  
 A L G P R G N S H S C D T V D G G Y Q C -  
  
 TTCCAGAAATTTCTCACTTGTGGGGTACCTACTCTCCATACTTCTCTTTGGCAGACGAA  
 121 -----+-----+-----+-----+-----+ 180  
 AAGGGTCTTTAAAGAGTGAACACCCCATGGATGAGAGGTATGAAGAGAAACCGTCTGCTT  
  
 F P E I S H L W G T Y S P Y F S L A D E -  
  
 TCTGCTATTTCTCCAGACGTCCCAAAGGACTGTAGAGTTACTTTTCGTTCAAGTTTTGTCT  
 181 -----+-----+-----+-----+-----+ 240  
 AGACGATAAAGAGGTCTGCAGGGTTTCTGACATCTCAATGAAAGCAAGTTCAAAACAGA  
  
 S A I S P D V P K D C R V T F V Q V L S -  
  
 AGACACGGTGCTAGATACCCAACTTCTTCTAAGTCTAAGGCTTACTCTGCTTTGATTGAA  
 241 -----+-----+-----+-----+-----+ 300  
 TCTGTGCCACGATCTATGGGTTGAAGAAGATTGAGATTCCGAATGAGACGAAACTAACTT  
  
 R H G A R Y P T S S K S K A Y S A L I E -  
  
 GCTATTCAAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAC  
 301 -----+-----+-----+-----+-----+ 360  
 CGATAAGTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTG  
  
 A I Q K N A T A F K G K Y A F L K T Y N -  
  
 TACACTTTGGGTGCTGACGACTTGACTCCATTTCGGTGAAAACCAAATGGTTAACTCTGGT  
 361 -----+-----+-----+-----+-----+ 420  
 ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACCAATTGAGACCA  
  
 Y T L G A D D L T P F G E N Q M V N S G -  
  
 ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCATTCTTAGAGCT  
 421 -----+-----+-----+-----+-----+ 480  
 TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA  
  
 I K F Y R R Y K A L A R K I V P F I R A -  
  
 TCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTTTCCAATCTGCT  
 481 -----+-----+-----+-----+-----+ 540  
 AGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACTTCCAAAGGTTAGACGA  
  
 S G S D R V I A S A E K F I E G F Q S A -

Fig. 22a

46/56

```

AAGTTGGCTGACCCAGGTTCTCAACCACACCAAGCTTCTCCAGTTATTAACGTGATCATT
541 -----+-----+-----+-----+-----+-----+ 600
TTCAACCGACTGGGTCCAAGAGTTGGTGTGGTTCGAAGAGGTCAATAATTGCACTAGTAA

K L A D P G S Q P H Q A S P V I N V I I -

CCAGAAGGATCCGGTTACAACAACACTTTGGACCATGGTCTTTGTACTGCTTTTGAAGAC
601 -----+-----+-----+-----+-----+-----+ 660
GGTCTTCCTAGGCCAATGTTGTTGTGAAACCTGGTACCAGAAACATGACGAAAGCTTCTG

P E G S G Y N N T L D H G L C T A F E D -

TCTACCCTAGGTGACGACGTTGAAGCTAACTTCACTGCTTTGTTTCGCTCCAGCTATTAGA
661 -----+-----+-----+-----+-----+-----+ 720
AGATGGGATCCACTGCTGCAACTTCGATTGAAGTGACGAAACAAGCGAGGTTCGATAATCT

S T L G D D V E A N F T A L F A P A I R -

GCTAGATTGGAAGCTGACTTGCCAGGTGTTACTTTGACTGACGAAGACGTTGTTTACTTG
721 -----+-----+-----+-----+-----+-----+ 780
CGATCTAACCTTCGACTGAACGGTCCACAATGAAACTGACTGCTTCTGCAACAAATGAAC

A R L E A D L P G V T L T D E D V V Y L -

ATGGACATGTGTCCATTTCGACACTGTGCTAGAACTTCTGACGCTACTGAATTGTCTCCA
781 -----+-----+-----+-----+-----+-----+ 840
TACCTGTACACAGGTAAGCTGTGACAGCGATCTTGAAGACTGCGATGACTTAACAGAGGT

M D M C P F D T V A R T S D A T E L S P -

TTCTGTGCTTTGTTCACTCACGACGAATGGATCCAATACGACTACTTGCAAAGCTTGGGT
841 -----+-----+-----+-----+-----+-----+ 900
AAGACACGAAACAAGTGAGTGCTGCTTACCTAGGTTATGCTGATGAACGTTTTCGAACCCA

F C A L F T H D E W I Q Y D Y L Q S L G -

AAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTGTTGGTTTCGCT
901 -----+-----+-----+-----+-----+-----+ 960
TTCATGATGCCAATGCCACGACCATTGGGTAAACCCAGGTCGAGTTCACAAACCAAAGCGA

K Y Y G Y G A G N P L G P A Q G V G F A -

AACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTTCTACTAACCAC
961 -----+-----+-----+-----+-----+-----+ 1020
TTGCTTAACTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGTGAAGATGATTGGTG

N E L I A R L T H S P V Q D H T S T N H -

ACTTTGGACTCTAACCCAGCTACTTTCCCATTTGAACGCTACTTTGTACGCTGACTTCTCT
1021 -----+-----+-----+-----+-----+-----+ 1080
TGAAACCTGAGATTGGGTTCGATGAAAGGGTAACTTGCGATGAAACATGCGACTGAAGAGA

T L D S N P A T F P L N A T L Y A D F S -

```

Fig. 22b

CCCTA" 5938450

47/56

1081 CACGACAACACTATGATATCTATTTTCTTCGCTTTGGGTTTGTACAACGGTACCAAGCCA  
-----+-----+-----+-----+-----+-----+ 1140  
GTGCTGTTGTGATACTATAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGGTTCGGT  
  
H D N T M I S I F F A L G L Y N G T K P -  
  
TTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACTCTGCTTCTTGACT  
1141 -----+-----+-----+-----+-----+-----+ 1200  
AACAGATGATGAAGACAACCTTAGATAAATTCTTTGACTGCCAATGAGACGAAGAACCTGA  
  
L S T T S V E S I E E T D G Y S A S W T -  
  
GTTCCATTTCGCTGCTAGAGCTTACGTTGAAATGATGCAATGTCAAGCTGAAAAGGAACCA  
1201 -----+-----+-----+-----+-----+-----+ 1260  
CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACAGTTCGACTTTTCCTTGGT  
  
V P F A A R A Y V E M M Q C Q A E K E P -  
  
TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGCTGTTGACAAG  
1261 -----+-----+-----+-----+-----+-----+ 1320  
AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACGACAACCTGTTTC  
  
L V R V L V N D R V V P L H G C A V D K -  
  
TTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTTCGCTAGATCTGGTGGT  
1321 -----+-----+-----+-----+-----+-----+ 1380  
AACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGATCTAGACCACCA  
  
L G R C K R D D F V E G L S F A R S G G -  
  
AACTGGGCTGAATGTTTCGCTTAA  
1381 -----+-----+----- 1404  
TTGACCCGACTTACAAAGCGAATT  
  
N W A E C F A \*

Fig. 22c

ATGGGCGTGTTTCGTGCTACTGTCCATTGCCACCTTGTTTCGGTTCACATCCGGTACC  
 1 -----+-----+-----+-----+-----+ 60  
 TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG  
  
 M G V F V V L L S I A T L F G S T S G T -  
  
 GCCTTGGGTCCTCGTGTAATTCTCACTCTTGTGACACTGTTGACGGTGGTTACCAATGT  
 61 -----+-----+-----+-----+-----+ 120  
 CGGAACCCAGGAGCACCATTAAGAGTGAGAACACTGTGACAACTGCCACCAATGGTTACA  
  
 A L G P R G N S H S C D T V D G G Y Q C -  
  
 TTCCCAGAAATTTCTCACTTGTGGGGTACCTACTCTCCATACTTCTCTTTGGCAGACGAA  
 121 -----+-----+-----+-----+-----+ 180  
 AAGGGTCTTTAAAGAGTGAACACCCCATGGATGAGAGGTATGAAGAGAAACCGTCTGCTT  
  
 F P E I S H L W G T Y S P Y F S L A D E -  
  
 TCTGCTATTTCTCCAGACGTCCCAAAGGACTGTAGAGTTACTTTTCGTTCAAGTTTTGTCT  
 181 -----+-----+-----+-----+-----+ 240  
 AGACGATAAAGAGGTCTGCAGGGTTTCCTGACATCTCAATGAAAGCAAGTTCAAAACAGA  
  
 S A I S P D V P K D C R V T F V Q V L S -  
  
 AGACACGGTGCTAGATACCCAACCTTCTTCTGCGTCTAAGGCTTACTCTGCTTTGATTGAA  
 241 -----+-----+-----+-----+-----+ 300  
 TCTGTGCCACGATCTATGGGTTGAAGAAGACGCAGATTCCGAATGAGACGAACTAATT  
  
 R H G A R Y P T S S A S K A Y S A L I E -  
  
 GCTATTCAAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAC  
 301 -----+-----+-----+-----+-----+ 360  
 CGATAAGTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTG  
  
 A I Q K N A T A F K G K Y A F L K T Y N -  
  
 TACACTTTGGGTGCTGACGACTTGACTCCATTTCGGTGAAAACCAAATGGTTAACTCTGGT  
 361 -----+-----+-----+-----+-----+ 420  
 ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACCAATTGAGACCA  
  
 Y T L G A D D L T P F G E N Q M V N S G -  
  
 ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCATTCAATTAGAGCT  
 421 -----+-----+-----+-----+-----+ 480  
 TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA  
  
 I K F Y R R Y K A L A R K I V P F I R A -  
  
 TCTGGTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTTTCCAATCTGCT  
 481 -----+-----+-----+-----+-----+ 540  
 AGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAAGTTCCAAAGGTTAGACGA  
  
 S G S D R V I A S A E K F I E G F Q S A -

Fig. 23a

[illegible][illegible]

50/56

CACGACAACACTATGATATCTATTTTCTTCGCTTTGGGTTTGTACAACGGTACCAAGCCA  
1081 -----+-----+-----+-----+-----+-----+ 1140  
GTGCTGTTGTGATACTATAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGGTTCGGT  
  
H D N T M I S I F F A L G L Y N G T K P -  
  
TTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACTCTGCTTCTTGACT  
1141 -----+-----+-----+-----+-----+-----+ 1200  
AACAGATGATGAAGACAACCTTAGATAACTTCTTTGACTGCCAATGAGACGAAGAACCTGA  
  
L S T T S V E S I E E T D G Y S A S W T -  
  
GTTCCATTTCGCTGCTAGAGCTTACGTTGAAATGATGCAATGTCAAGCTGAAAAGGAACCA  
1201 -----+-----+-----+-----+-----+-----+ 1260  
CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACAGTTCGACTTTTCCTTGGT  
  
V P F A A R A Y V E M M Q C Q A E K E P -  
  
TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGCTGTTGACAAG  
1261 -----+-----+-----+-----+-----+-----+ 1320  
AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACGACAACCTGTTTC  
  
L V R V L V N D R V V P L H G C A V D K -  
  
TTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTTCGCTAGATCTGGTGGT  
1321 -----+-----+-----+-----+-----+-----+ 1380  
AACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGATCTAGACCACCA  
  
L G R C K R D D F V E G L S F A R S G G -  
  
AACTGGGCTGAATGTTTCGCTTAA  
1381 -----+-----+----- 1404  
TTGACCCGACTTACAAAGCGAATT  
  
N W A E C F A \*

Fig. 23c

ATGGGCGTGTTCGTGCTGCTACTGTCCATTGCCACCTTGTTCGGTTCACATCCGGTACC  
 1 -----+-----+-----+-----+-----+-----+ 60  
 TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG  
  
 M G V F V V L L S I A T L F G S T S G T -  
  
 GCCTTGGGTCCTCGTGGTAATTCTCACTCTTGTGACACTGTTGACGGTGGTTACCAATGT  
 61 -----+-----+-----+-----+-----+-----+ 120  
 CGGAACCCAGGAGCACCATTAAAGAGTGAGAACACTGTGACAACCTGCCACCAATGGTTACA  
  
 A L G P R G N S H S C D T V D G G Y Q C -  
  
 TTCCCAGAAATTTCTCACTTGTGGGGTACATACTCTCCATTCTTCTCTTTGGCTGACGAA  
 121 -----+-----+-----+-----+-----+-----+ 180  
 AAGGGTCTTTAAAGAGTGAACACCCCATGTATGAGAGGTAAGAAGAGAAACCGACTGCTT  
  
 F P E I S H L W G T Y S P F F S L A D E -  
  
 TCTGCTATTCTCCAGACGTTCCAAAGGGTGTAGAGTTACTTTGTTCAAGTTTTGTCT  
 181 -----+-----+-----+-----+-----+-----+ 240  
 AGACGATAAAGAGGTCTGCAAGGTTTCCCAACATCTCAATGAAAGCAAGTTCAAAACAGA  
  
 S A I S P D V P K G C R V T F V Q V L S -  
  
 AGACACGGTGCTAGATACCCAACCTTCTTCTAAGTCTAAGGCTTACTCTGCTTTGATTGAA  
 241 -----+-----+-----+-----+-----+-----+ 300  
 TCTGTGCCACGATCTATGGGTTGAAGAAGATTGAGATTCCGAATGAGACGAACTAATCTT  
  
 R H G A R Y P T S S K S K A Y S A L I E -  
  
 GCTATTCAAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAT  
 301 -----+-----+-----+-----+-----+-----+ 360  
 CGATAAGTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTA  
  
 A I Q K N A T A F K G K Y A F L K T Y N -  
  
 TACACTTTGGGTGCTGACGACTTGACTCCATTCCGGTGAACAACAAATGGTTAACTCTGGT  
 361 -----+-----+-----+-----+-----+-----+ 420  
 ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTGTGTGTTACCAATTGAGACCA  
  
 Y T L G A D D L T P F G E Q Q M V N S G -  
  
 ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCATTCTTAGAGCT  
 421 -----+-----+-----+-----+-----+-----+ 480  
 TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA  
  
 I K F Y R R Y K A L A R K I V P F I R A -  
  
 TCTGGTTCTGACAGAGTTATTGCTTCTGCCGAAAAGTTCATTGAAGGTTTCCAATCTGCT  
 481 -----+-----+-----+-----+-----+-----+ 540  
 AGACCAAGACTGTCTCAATAACGAAGACGGCTTTTCAAGTAACTTCCAAAGGTTAGACGA  
  
 S G S D R V I A S A E K F I E G F Q S A -

Fig. 24a

Fig. 24b

53/56

```

CACGACAACACTATGGTTTCTATTTTCTTCGCTTTGGGTTTGTACAACGGTACTAAGCCA
1081 -----+-----+-----+-----+-----+ 1140
GTGCTGTTGTGATACCAAAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGATTTCGGT

H D N T M V S I F F A L G L Y N G T K P -

TTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACTCTGCTTCTTGGA
1141 -----+-----+-----+-----+-----+ 1200
AACAGATGATGAAGACAACCTTAGATAAATTCTTTGACTGCCAATGAGACGAAGAACCTGA

L S T T S V E S I E E T D G Y S A S W T -

GTTCCATTTCGCTGCTAGAGCTTACGTTGAAATGATGCAATGTGAAGCTGAAAAGGAACCA
1201 -----+-----+-----+-----+-----+ 1260
CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACACTTCGACTTTTCCTTGGT

V P F A A R A Y V E M M Q C E A E K E P -

TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGCTGTTGACAAG
1261 -----+-----+-----+-----+-----+ 1320
AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACGACAACCTGTTT

L V R V L V N D R V V P L H G C A V D K -

TTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTTCGCTAGATCTGGTGGT
1321 -----+-----+-----+-----+-----+ 1380
AACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGATCTAGACCACCA

L G R C K R D D F V E G L S F A R S G G -

AACTGGGAAGAATGTTTCGCTTAA
1381 -----+-----+----- 1404
TTGACCCTTCTTACAAAGCGAATT

N W E E C F A *
```

Fig. 24c

ATGGGCGTGTTTCGTGCTACTGTCCATTGCCACCTTGTTTCGGTTCACATCCGGTACC  
 1 -----+-----+-----+-----+-----+-----+ 60  
 TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG  
  
 M G V F V V L L S I A T L F G S T S G T -  
  
 GCCTTGGGTCCTCGTGGTAATTCTCACTCTTGTGACACTGTTGACGGTGGTTACCAATGT  
 61 -----+-----+-----+-----+-----+-----+ 120  
 CGGAACCCAGGAGCACCATTAAGAGTGAGAACACTGTGACAACCTGCCACCAATGGTTACA  
  
 A L G P R G N S H S C D T V D G G Y Q C -  
  
 TTCCCAGAAATTTCTCACTTGTGGGGTACATACTCTCCATTCTTCTCTTTGGCTGACGAA  
 121 -----+-----+-----+-----+-----+-----+ 180  
 AAGGGTCTTTAAAGAGTGAACACCCCATGTATGAGAGGTAAGAAGAGAAACCGACTGCTT  
  
 F P E I S H L W G T Y S P F F S L A D E -  
  
 TCTGCTATTCTCCAGACGTTCCAAAGGGTTGTAGAGTTACTTTTCGTTCAAGTTTTGTCT  
 181 -----+-----+-----+-----+-----+-----+ 240  
 AGACGATAAAGAGGTCTGCAAGGTTTCCCAACATCTCAATGAAAGCAAGTTCAAAACAGA  
  
 S A I S P D V P K G C R V T F V Q V L S -  
  
 AGACACGGTGCTAGATACCCAACTTCTTCTGCGTCTAAGGCTTACTCTGCTTTGATTGAA  
 241 -----+-----+-----+-----+-----+-----+ 300  
 TCTGTGCCACGATCTATGGGTTGAAGAAGACGCAGATTCCGAATGAGACGAAACTAACTT  
  
 R H G A R Y P T S S A S K A Y S A L I E -  
  
 GCTATTCAAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAT  
 301 -----+-----+-----+-----+-----+-----+ 360  
 CGATAAGTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTA  
  
 A I Q K N A T A F K G K Y A F L K T Y N -  
  
 TACACTTTGGGTGCTGACGACTTGACTCCATTTCGGTGAACAACAAATGGTTAACTCTGGT  
 361 -----+-----+-----+-----+-----+-----+ 420  
 ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTGTTGTTTACCAATTGAGACCA  
  
 Y T L G A D D L T P F G E Q Q M V N S G -  
  
 ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCATTCTTAGAGCT  
 421 -----+-----+-----+-----+-----+-----+ 480  
 TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA  
  
 I K F Y R R Y K A L A R K I V P F I R A -  
  
 TCTGGTTCTGACAGAGTTATTGCTTCTGCCGAAAAGTTCATTGAAGGTTTCCAATCTGCT  
 481 -----+-----+-----+-----+-----+-----+ 540  
 AGACCAAGACTGTCTCAATAACGAAGACGGCTTTTCAAGTAACTTCCAAAGGTTAGACGA  
  
 S G S D R V I A S A E K F I E G F Q S A -

Fig. 25a

## 55/56

AAGTTGGCTGACCCAGGTGCTAACCCACACCAAGCTTCTCCAGTTATTAACGTTATTATT  
 541 -----+-----+-----+-----+-----+ 600  
 TTCAACCGACTGGGTCCACGATTGGGTGTGGTTCGAAGAGGTCAATAATTGCAATAATAA  
  
 K L A D P G A N P H Q A S P V I N V I I -  
  
 CCAGAAGGTGCTGGTTACAACAACACTTTGGACCACGGTTTGTGTACTGCTTTCGAAGAA  
 601 -----+-----+-----+-----+-----+ 660  
 GGTCTTCCACGACCAATGTTGTTGTGAAACCTGGTGCCAAACACATGACGAAAGCTTCTT  
  
 P E G A G Y N N T L D H G L C T A F E E -  
  
 TCTACCCTAGGTGACGACGTTGAAGCTAACTTCACTGCTGTTTTTCGCTCCACCAATTAGA  
 661 -----+-----+-----+-----+-----+ 720  
 AGATGGGATCCACTGCTGCAACTTCGATTGAAGTGACGACAAAAGCGAGGTGGTTAATCT  
  
 S T L G D D V E A N F T A V F A P P I R -  
  
 GCTAGATTGGAAGCTCACTTGCCAGGTGTTAACTTGACTGACGAAGACGTTGTTAACTTG  
 721 -----+-----+-----+-----+-----+ 780  
 CGATCTAACCTTCGAGTGAACGGTCCACAATTGAACTGACTGCTTCTGCAACAATTGAAC  
  
 A R L E A H L P G V N L T D E D V V N L -  
  
 ATGGACATGTGTCCATTCGACACTGTTGCTAGAAGTTCTGACGCTACTCAATTGTCTCCA  
 781 -----+-----+-----+-----+-----+ 840  
 TACCTGTACACAGGTAAGCTGTGACAACGATCTTGAAGACTGCGATGAGTTAACAGAGGT  
  
 M D M C P F D T V A R T S D A T Q L S P -  
  
 TTCTGTGACTTGTTCCTCAGCAGCAATGGATTCAATACGACTACTTGCAATCTTTGGGT  
 841 -----+-----+-----+-----+-----+ 900  
 AAGACACTGAACAAGTGAGTGCTGCTTACCTAAGTTATGCTGATGAACGTTAGAAACCCA  
  
 F C D L F T H D E W I Q Y D Y L Q S L G -  
  
 AAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTGTTGGTTTCGTT  
 901 -----+-----+-----+-----+-----+ 960  
 TTCATGATGCCAATGCCACGACCATTTGGGTAACCCAGGTCGAGTTCACAACCAAAGCAA  
  
 K Y Y G Y G A G N P L G P A Q G V G F V -  
  
 AACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTTCTACTAACCAC  
 961 -----+-----+-----+-----+-----+ 1020  
 TTGCTTAACTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGTGAAGATGATTGGTG  
  
 N E L I A R L T H S P V Q D H T S T N H -  
  
 ACTTTGGACTCTAACCCAGCTACTTTCCCATTTGAACGCTACTTTGTACGCTGACTTCTCT  
 1021 -----+-----+-----+-----+-----+ 1080  
 TGAAACCTGAGATTGGGTGATGAAAGGGTAACTTGCGATGAAACATGCGACTGAAGAGA  
  
 T L D S N P A T F P L N A T L Y A D F S -

Fig. 25b

56/56

CACGACAACACTATGGTTTCTATTTTCTTCGCTTTGGGTTTGTACAACGGTACTAAGCCA  
1081 -----+-----+-----+-----+-----+-----+ 1140  
GTGCTGTTGTGATACCAAAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGATTCCGGT  
  
H D N T M V S I F F A L G L Y N G T K P -  
  
TTGTCTACTACTTCTGTTGAATCTATTGAAGAACTGACGGTTACTCTGCTTCTTGGA  
1141 -----+-----+-----+-----+-----+-----+ 1200  
AACAGATGATGAAGACAACTTAGATAACTTCTTTGACTGCCAATGAGACGAAGAACCCTGA  
  
L S T T S V E S I E E T D G Y S A S W T -  
  
GTTCCATTTCGCTGCTAGAGCTTACGTTGAAATGATGCAATGTGAAGCTGAAAAGGAACCA  
1201 -----+-----+-----+-----+-----+-----+ 1260  
CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACACTTCGACTTTTCCTTGGT  
  
V P F A A R A Y V E M M Q C E A E K E P -  
  
TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGCTGTTGACAAG  
1261 -----+-----+-----+-----+-----+-----+ 1320  
AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACGACAACCTGTTTC  
  
L V R V L V N D R V V P L H G C A V D K -  
  
TTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTTCGCTAGATCTGGTGGT  
1321 -----+-----+-----+-----+-----+-----+ 1380  
AACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGATCTAGACCACCA  
  
L G R C K R D D F V E G L S F A R S G G -  
  
AACTGGGAAGAATGTTTCGCTTAA  
1381 -----+-----+----- 1404  
TTGACCCTTCTTACAAAGCGAATT  
  
N W E E C F A \*

Fig. 25c

**COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY**  
(Includes Reference to PCT International Applications)

Attorney's Docket Number:  
**5808.200-US**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Improved Phytases

The specification of which (check only one item below):

☐ is attached hereto

☒ was filed as United States application

Application No. To Be Assigned

on January 20, 2000

and was amended

on \_\_\_\_\_

☐ was filed as PCT international application  
Number \_\_\_\_\_

on \_\_\_\_\_

and was amended under PCT Article 19

on \_\_\_\_\_

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by an amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability of this application in accordance with Title 37, Code of Federal Regulations, §1.56.

I hereby claim priority benefits under Title 35, United States Code, §119 of any provisional or foreign application(s) for patent or inventor's certificate or of any PCT international applications(s) for patent or inventor's certificate or of any PCT international applications(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

**PRIOR U.S. PROVISIONAL/FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. 119:**

| COUNTRY<br>(if PCT, indicated "PCT") | APPLICATION NUMBER | DATE OF FILING<br>(day, month, year) | PRIORITY CLAIMED<br>UNDER 35 USC 119                                |
|--------------------------------------|--------------------|--------------------------------------|---------------------------------------------------------------------|
| Denmark                              | PA 1999 00092      | 22 January 1999                      | <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO |
| Denmark                              | PA 1999 01340      | 21 September 1999                    | <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO |
| United States                        | 60/117,659         | 28 January 1999                      | <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO |
| United States                        | 60/156,495         | 28 September 1999                    | <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO |
|                                      |                    |                                      | <input type="checkbox"/> YES <input type="checkbox"/> NO            |

## 5808.200-US

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT  
UNDER 35 U.S.C. 120:

| U S APPLICATIONS                      |                 |                                        | STATUS (Check one) |         |           |
|---------------------------------------|-----------------|----------------------------------------|--------------------|---------|-----------|
| U S APPLICATION NUMBER                | U S FILING DATE |                                        | Patented           | Pending | Abandoned |
|                                       |                 |                                        |                    |         |           |
|                                       |                 |                                        |                    |         |           |
|                                       |                 |                                        |                    |         |           |
| PCT APPLICATIONS DESIGNATING THE U.S. |                 |                                        |                    |         |           |
| APPLICATION NO                        | FILING DATE     | US SERIAL NUMBERS<br>ASSIGNED (if any) |                    |         |           |
|                                       |                 |                                        |                    |         |           |
|                                       |                 |                                        |                    |         |           |
|                                       |                 |                                        |                    |         |           |

**Send Correspondence to:** Steve T. Zelson, Esq.  
Novo Nordisk of North America, Inc.  
405 Lexington Avenue, Suite 6400  
New York, New York 10174-6400

**Direct Telephone Calls To:**  
Steve T Zelson  
(212) 867-0123

|   |                         |                                               |                                                  |                                                          |
|---|-------------------------|-----------------------------------------------|--------------------------------------------------|----------------------------------------------------------|
| 1 | Full Name of Inventor   | Family Name<br><b>Lehmann</b>                 | First Given Name<br><b>Martin</b>                | Second Given Name                                        |
|   | Residence & Citizenship | City<br><b>Princeton</b>                      | State or Foreign Country<br><b>United States</b> | Country of Citizenship<br><b>Germany</b>                 |
|   | Post Office Address     | Post Office Address<br><b>258 Sayre Drive</b> | City<br><b>Princeton</b>                         | State & Zip Code/Country<br><b>New Jersey 07043, USA</b> |
| 2 | Full Name of Inventor   | Family Name                                   | First Given Name                                 | Second Given Name                                        |
|   | Residence & Citizenship | City                                          | State or Foreign Country                         | Country of Citizenship                                   |
|   | Post Office Address     | Post Office Address                           | City                                             | State & Zip Code/Country                                 |
| 3 | Full Name of Inventor   | Family Name                                   | First Given Name                                 | Second Given Name                                        |
|   | Residence & Citizenship | City                                          | State or Foreign Country                         | Country of Citizenship                                   |
|   | Post Office Address     | Post Office Address                           | City                                             | State & Zip Code/Country                                 |
| 4 | Full Name of Inventor   | Family Name                                   | First Given Name                                 | Second Given Name                                        |
|   | Residence & Citizenship | City                                          | State or Foreign Country                         | Country of Citizenship                                   |
|   | Post Office Address     | Post Office Address                           | City                                             | State & Zip Code/Country                                 |

[illegible]

## 5808.200-US

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon

Date \_\_\_\_\_